NBER WORKING PAPER SERIES

OPTIMAL MONEY BURNING: THEORY AND APPLICATION TO CORPORATE DIVIDEND POLICY

B. Douglas Bernheim
Lee Redding

NBER Working Paper 5682

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 1996

The work reported here was supported by the National Science Foundation through grants SES-9110211 and SES-8821666. We are grateful to Raymond Wolff and Jin Park for assistance with data processing. John Campbell, Tim Vogelsang, and an anonymous referee provided helpful comments. This paper is part of NBER's research program in Public Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper 5682
July 1996

OPTIMAL MONEY BURNING: THEORY AND APPLICATION TO CORPORATE DIVIDEND POLICY

ABSTRACT

We explore signaling behavior in settings with a discriminating signal and several costly nondiscriminating ("money burning") activities. In settings where informed parties have many options for burning money, existing theory provides no basis for selecting one nondiscriminating activity over another. When senders have private information about the costs of these activities, each sender's indifference is resolved, the taxation of a nondiscriminating signal is Pareto improving, and the use of the taxed activity becomes more widespread as the tax rate rises. We apply this analysis to the theory of dividend signaling. The central testable implication of the model is verified empirically.

B. Douglas Bernheim
Department of Economics
Stanford University
Stanford, CA 94305-6072
and NBER

Lee Redding
Department of Economics
Fordham University
441 East Fordham Road
Bronx, NY 10458-5158

I. Introduction

Models involving the signaling of private information through costly activities have been used to explain a wide range of economic phenomena. Signaling in the style of Spence [1974] requires the existence of a "discriminating" activity which is more costly for low quality types than for high quality types. When the *marginal* cost of this activity is always higher for low quality types than for high quality types, the Spence-Mirrlees "single crossing property" is said to hold. If the single crossing property fails, it may still be possible to support a signaling equilibrium through the use of a discriminating activity (as long as *total* costs are higher for the low quality type); however, this may involve a level of activity at which marginal costs are higher for high quality types than for low quality types. In that case, it is in the interest of high quality types to differentiate themselves by using the discriminatory activity (at a lower level) in combination with an activity that is "nondiscriminating," in the sense that its marginal cost is the same irrespective of quality. This observation gives rise to theories in which economic agents have incentives to engage in observable, nondiscriminating resource dissipation, or "money burning."

As we discuss below, previous signaling models have usually posited at most one mechanism for burning money. In most settings, however, there are many ways to dissipate resources observably, and existing theory provides no basis for selecting one nondiscriminatory signal over another. This is sometimes viewed as an inherent shortcoming of all signaling-cum-money-burning explanations of economic behavior.

For concreteness, consider the analysis of Milgrom and Roberts [1986], who present a model in which a firm wishes to signal that its products are of high quality. Managers of firms use two choice variables to convey information to the public: the introductory price of the good, and the amount spent on advertising. The costs of advertising are assumed to be independent of type. In contrast, the costs and benefits of varying introductory price do depend on the firm's unobserved characteristics. Thus, varying introductory price is a discriminatory activity, while advertising is a nondiscriminatory activity. Since the single crossing property does not hold for variations in introductory price, burning money through advertising may emerge as part of an efficient equilibrium signal.

In a related model, Bagwell and Bernheim [1996] consider conspicuous consumption as a means of signaling personal wealth. Varying expenditures on a conspicuous good discriminates between individuals on the basis of wealth over some ranges. In contrast, reducing consumption of the conspicuous good for any fixed level of expenditure (which is accomplished by overpaying for the conspicuous good) is a nondiscriminatory activity. When the single crossing property fails to hold globally for total expenditures, Veblen effects (defined as a preference to pay more for a functionally equivalent branded good) may therefore arise as part of an efficient signal.

In models such as these, the choice of the means to burn money is inherently arbitrary.[1] Advertising and conspicuous consumption are chosen as means to burn money primarily on the basis of their visibility. Yet other methods of burning money are plainly available. For example, in the context of Milgrom and Roberts' analysis, corporate donations to charity could in principle fill the same role. Generally, the question of why one method of money burning is chosen above all others has been left open. This paper presents a model, based on a suggestion in Bernheim [1991], that explores this choice on a more formal basis.

The motivation and intuition for the current analysis can be understood as follows. It is important in all money-burning models that nondiscriminating activities dissipate resources at a known rate that is common for all types. However, with many potential money-burning activities, the true rate of resource dissipation for the sender is not necessarily known to receivers. Consider, for example, signaling of product quality by corporations. Advertising may confer some benefit on the firm in terms of image or public relations. The true amount of money burned is advertising expenses net of these benefits. Moreover, managers may have private information about the size of these benefits, even when advertising is nondiscriminatory (in the sense that net benefits are unrelated to quality). Likewise, managers may receive "psychic income" from their corporations' charitable contributions, with the magnitude of this benefit varying in ways that are both unrelated to product quality, and known only to the manager. As private information about the true rate of resource dissipation becomes more important, the problem of inference may also become more problematic. For example, if one observes a company making substantial charitable contributions, does one infer that the company must produce a high quality product, or that the managers have relatively strong preferences for charity?

In this paper, we explore the optimal method of burning money in situations where several nondiscriminating activities are available as potential signals (in addition to a single discriminating activity). The incentive to burn money arises from our assumption that the single discriminating activity fails to satisfy the single crossing property globally. The true net cost of each nondiscriminatory activity is agent-specific and known only to the sender, but is unrelated to the sender's quality. We demonstrate that, in this setting, each agent strictly prefers a single money-burning activity to all others. However, since the preferred form of money burning differs from agent to agent, the theory generates an endogenous distribution of signaling strategies. In general, every potential money-burning activity will be chosen as a method of signaling by some positive fraction of agents.

This theory of optimal money burning has the following two strong, and rather surprising implications: first, an increase in the observable component of the cost of a nondiscriminatory activity (e.g. resulting from an increase in the tax rate on that activity) leads to *more* widespread use of that activity as a money-burning signal; second, the taxation of a nondiscriminating signal is Pareto improving. Intuitively, an increase in an observable tax on an activity reduces the significance of private information about the level of the activity required to dissipate the equivalent of one dollar, thereby "purifying" the activity as a signal of quality. It is worth mentioning that once an agent begins to use an activity as a money-burning signal, further increases in the tax rate reduce the level of the activity undertaken in equilibrium. Thus, an increase in taxation has an ambiguous effect on the aggregate volume of the activity.

These general results are then applied to the case of a corporation signaling financial strength to investors through the use of cash distributions to shareholders. Since the tax code treats cash dividends unfavorably relative to other methods of increasing net cash flows to shareholders (particularly when the firm is also issuing new equity), the practice of paying dividends is widely viewed as a puzzle. Bernheim [1991] points out that, in the corporate setting, limited liability naturally leads to a failure of the single crossing property, and that the optimal signal may therefore combine net distributions with a nondiscriminatory money burning activity, such as making net distributions in a tax-disfavored form. Naturally, firms may also choose other methods of burning money, such as charity, unproductive advertising, and wanton destruction of plant and equipment. However, Bernheim argues that dividends offer a better method of nondiscriminating dissipation for most firms because the investing public has better information about the associated tax costs (which, after all, are incurred by the investors rather than by the firm).

The current analysis demonstrates that Bernheim's theory has the following strong implications: the taxation of dividends is Pareto improving, and an increase in the relative tax burden on dividends should increase the number of firms that use dividends to signal profitability (it may increase or decrease the total volume of dividends). The second implication is potentially testable. Although previous studies have found a negative relation between dividend taxation and the aggregate level of dividends, this is consistent with our theory, and does not rule out the possibility that the tax might nevertheless increase the number of firms choosing to pay dividends. We explore this latter possibility empirically, and find that there is indeed a reasonably strong tendency for the fraction of corporations that pay dividends to move in the same direction as the dividend tax rate, even when other factors are accounted for.

The paper proceeds as follows. Section II outlines the general signaling model and derives results. Section III applies the theory to the case of firms paying cash dividends to signal high profitability. It also contains empirical tests of the central hypothesis. Section IV concludes.

II. Theory

A. The Model

We consider a signaling model consisting of a "sender" and a "receiver." The sender has private information concerning a vector of characteristics, Ω. One can think of Ω as the sender's "type." We decompose type as follows: $\Omega = (Q,\gamma)$, where Q will be interpreted as the sender's "quality," while γ will summarize other characteristics, detailed below. We will assume for simplicity that quality is either "high" (Q_H) or "low" (Q_L). The sender initiates the process of interaction by choosing a "message," $m = (x,z)$, where x (a scalar) will be a potentially discriminating activity, and z (a K-dimensional vector) will consist of nondiscriminating activities. Each component of the message m is constrained to be non-negative.

Having observed the sender's message, the receiver selects a "response," R. The receiver's response may depend upon inferences about the sender's type. Thus, we write the receiver's response as $R(\rho)$, where ρ represents the receiver's probabilistic assessment that the sender is of high quality, given the observed message m. Henceforth, we will use R_H to denote R(1), and R_L to denote R(0). We assume that $R_H \neq R_L$, and we adopt the normalization that $R_H > R_L$.

In the preceding paragraph, we have intentionally been vague about the process that determines the response function $R(\rho)$. In some instances, we will impose the following assumption:

Assumption R

> $R(\rho)$ maximizes $\rho u(R,Q_H) + (1-\rho)u(R,Q_L)$, where u(R,Q) represents the receiver's payoff associated with response R and quality Q.

Under assumption R, the receiver does not care directly about x or z. However, these variables may convey information about the sender's type, and therefore may affect the receiver's response. One can, of course, imagine more complex processes through which the response might be determined. For example, in many kinds of signaling models, the response is influenced by competition among alternative receivers. Consequently, in most instances, we will refrain from imposing assumption R, and instead allow for more general processes.

The sender's payoff is given by $U(x,z,R,\Omega)$. We assume the costs and benefits of x depend on Q and R, while the costs and benefits of z depend only on γ. Thus, only x potentially discriminates between possible senders on the basis of quality. We will also assume that γ consists of a K-dimensional vector, the i-th component of which measures the constant marginal cost of the i-th non-quality-discriminating activity. More specifically, the sender's utility is given by

$$U(x,z,R,\Omega) = V(x,R,Q) - \sum_{k=1}^{K} \gamma_k z_k , \tag{1}$$

where V is differentiable, strictly decreasing in x, and strictly increasing in R (so that high quality senders have an incentive to communicate their private information about Q to the receiver).

For signaling to be feasible in this setting, the activity x must be more costly for low quality senders than for high quality senders, at least over some range. We do not, however, wish to assume that the single crossing property is satisfied globally, since this would imply that it is inefficient to signal by burning money (i.e. by undertaking a combination of x and z). Instead, we assume that the marginal cost of x is higher for lower quality senders only over some range: $V_x(x,R,Q_L) < V_x(x,R,Q_H)$ iff $0 < x < \hat{x}(R)$, where $\hat{x}(R)$ is positive for all R.

The single crossing property is equivalent to the assumption that $\hat{x}(R) = +\infty$.

We will use π to denote the probability that a sender's quality is Q_H. We will assume that Q is distributed independently of γ. Each marginal cost parameter γ_k may take on any value in the interval $[L_k,B_k]$, where $L_k > 0$ (so that each activity definitely has a positive cost), and where B_k may be infinite. The probability density function describing the distribution of the vector γ is denoted $f(\gamma_1,\ldots,\gamma_K)$. We will assume that the support of this distribution is $\Gamma \equiv \times_{k=1}^{K} [L_k,B_k]$, so that all possible types are potentially represented. Finally, both π and f are taken to be public knowledge.

We study equilibria in which potential senders achieve full separation of types along the dimension of quality, Q.[2] As in most signaling contexts, our model can give rise to a variety of separating equilibria.[3] We take the view that the requirements of perfect Bayesian equilibria (see Fudenberg and Tirole [1991]) are minimal restrictions on the reasonableness of an outcome, and we refine the equilibrium set accordingly. Using standard arguments, one can easily show that, in any perfect Bayesian separating equilibrium, low quality types choose m = 0.[4] Non-imitation by all low quality types then requires each high quality type to select (x,z) satisfying

(2) $$V(x,R_H,Q_L) - \sum_{k=1}^{K} \gamma_k z_k \leq V(0,R_L,Q_L)$$

for all $\gamma \in \Gamma$. This is, of course, equivalent to

(3) $$V(x,R_H,Q_L) - \sum_{k=1}^{K} L_k z_k \leq V(0,R_L,Q_L) \ .$$

We will further refine the set of separating equilibrium by assuming that each high quality type differentiates itself from lower quality types in the most efficient manner possible.

Formally, we focus on equilibria in which type (Q_H,γ) chooses (x,z) to maximize $U(x,z,R_H,Q_H,\gamma)$, subject to expression (3) and non-negativity constraints on x and z.[5] Using standard arguments, one can easily verify that this is the only separating equilibrium which survives the application of the "equilibrium dominance" refinement proposed by Cho and Kreps [1987].

B. The Choice of a Nondiscriminating Signal

Our primary focus in this paper is on the sender's choice of a nondiscriminating signal. The following result provides a characterization of this choice:

Theorem 1

The most efficient signal for agents of type (Q_H,γ) *involves the use of the* k-th *nondiscriminating activity if and only if* $\lambda_k \equiv \gamma_k/L_k \leq \gamma_j/L_j \equiv \lambda_j$ *for* $j \in \{1,2,\ldots,K\}$.

Proof. Suppose not. Then there are two activities, j and k, such that $\lambda_j > \lambda_k$, and such that $z_j^* > 0$ (where stars denote equilibrium values). Define the vector $\hat{z}$ as follows: $\hat{z}_i = z_i^*$ for each $i \neq j,k$; $\hat{z}_j = 0$; and $\hat{z}_k = z_k^* + (L_j/L_k)z_j^*$. Notice that, since (x^*,z^*) satisfies (3) and non-negativity, then $(x^*,\hat{z})$ also satisfies (3) and non-negativity. Moreover,

$$U(x^*,\hat{z},R_H,Q_H,\gamma) = U(x^*,z^*,R_H,Q_H,\gamma) + z_j^*\left[\gamma_j - \gamma_k\left(\frac{L_j}{L_k}\right)\right] \quad . \tag{4}$$

Under our supposition, the last term on the right hand side of (4) is strictly positive. Thus, (x^*,z^*) could not maximize $U(x,z,R_H,Q_H,\gamma)$ subject to (3) and non-negativity. ■

There are several important implications of this theorem. First, the model resolves the problem of indifference across nondiscriminating methods of burning money. Except on a set of measure zero, each agent strictly prefers one nondiscriminating activity to all others. Note that the resolution of this indifference depends critically on the existence of private information concerning rates of resource dissipation for nondiscriminating activities. In the absence of private information of this sort, the central indeterminacy would remain, even if rates of dissipation differed both across activities and across senders.

Second, different agents will use different nondiscriminating signals. Each type is most attracted to those activities for which his or her relative costs (compared to the lowest cost incurred by any type) are smallest. Thus, for example, managers with strong preferences for charity will be more likely to burn money by making charitable contributions than by advertising.

Third, as long as there is some private information about the costs of all activities, every nondiscriminating activity is used by some agents to burn money in equilibrium. This is because, for each activity j, some agents must have γ_j very close to L_j. In fact, it is possible to derive an explicit expression for the fraction of high quality senders choosing any particular activity. Without loss of generality, we write this expression for activity 1:

(5) $$\int_{L_1}^{B_1} \int_{M_2}^{B_2} \cdots \int_{M_K}^{B_K} f(\gamma_1, \gamma_2, \ldots, \gamma_K)\, d\gamma_K \ldots d\gamma_2\, d\gamma_1 \quad ,$$

where $M_j \equiv (L_j/L_1)\,\gamma_1$. A nondiscriminating activity is used by a larger fraction of the population when the distribution of marginal costs for that activity are concentrated near (in relative terms) the lowest marginal cost for that activity. This will tend to be the case when, for example, there is relatively little private information about the cost of an activity. In the limiting case where the costs of an activity are uniform throughout the population, all agents will (at least weakly) prefer to burn money using that activity. At the other extreme, there is a tendency not to burn money through activities for which there is a slight possibility that some agents have little or no costs.

C. Activity Levels

As noted above, the most efficient signal for type (Q_H,γ) can be derived by maximizing the type's utility subject to (3) and non-negativity. Since type (Q_H,γ) uses a single nondiscriminating activity, k (where $\lambda_k \leq \lambda_j$ for $j \in \{1,2,\ldots,K\}$), the efficient signal is also the solution to the following problem:

(6) $$\max_{x,z_k} V(x,R_H,Q_H) - \gamma_k z_k$$

subject to

(7) $$V(x,R_H,Q_L) - L_k z_k \leq V(0,R_L,Q_L) \quad ,$$

in addition to $x,\ z_k \geq 0$.

We will consider interior solutions to this problem (where the non-negativity constraint on z_k is non-binding). It is easy to verify that (7) (the incentive compatibility constraint) must hold with equality. By substituting the binding constraint into the objective function, we obtain the following equivalent problem:

(8) $$\max_x V(x,R_H,Q_H) - \lambda_k[V(x,R_H,Q_L) - V(0,R_L,Q_L)] \; .$$

The solution to this problem, denoted x^*, is characterized by the following first order condition:

(9) $$V_x(x^*,R_H,Q_H) = \lambda_k V_x(x^*,R_H,Q_L) \leq V_x(x^*,R_H,Q_L) \quad ,$$

where $V_x(x,R,Q)$ denotes the partial derivative of V with respect to x.[6] The equilibrium level of the nondiscriminating activity, z_k^*, is then chosen to satisfy (7):

(10) $$z_k^* = \left(\frac{1}{L_k}\right)\left[V(x,R_H,Q_L) - V(0,R_L,Q_L)\right] \; .$$

Equation (9) immediately establishes that money burning is never optimal when preferences respect the single crossing property (since the condition for an interior optimum could never be satisfied). However, if (as assumed in this paper), given R_H, there is some $\hat{x}(R_H)$ such that the marginal costs of x are inversely related to quality for $x < \hat{x}(R_H)$, but positively related to quality for $x > \hat{x}(R_H)$, then one can in principle have an interior optimum involving money burning. By equation (9), this assumption has the additional implication that the optimal signal involves a level of x in excess of $\hat{x}(R_H)$.

It is of interest to note that, while the equilibrium achieves full separation along the quality dimension, it also fully reveals γ_k. In particular, each type choosing to burn money through activity k will select a different level of x and z_k. This follows from implicit differentiation of (9):

$$(11) \qquad \frac{dx^*}{d\gamma_k} = \frac{V_x(x^*,R_H,Q_L)/L_k}{V_{xx}(x^*,R_H,Q_H) - \lambda_k V_{xx}(x^*,R_H,Q_L)} > 0$$

(where we have used the second order condition to sign the denominator). Note in particular that high quality agents will rely more heavily on x, and less heavily on z_k, when their individual costs of z_k are higher.

D. The Effects of Taxation

We now consider the effects of taxing a nondiscriminating activity. Let k denote the index of the activity that is subject to taxation. We will suppose throughout the following analysis that, fixing x, z, and R, the tax reduces the utility of every agent, regardless of type, by the amount τ for each unit of z_k. This is easily incorporated into the preceding analysis by defining $L_k \equiv L_k^o + \tau$, $B_k \equiv B_k^o + \tau$, and, for each sender, $\gamma_k = \gamma_k^o + \tau$. That is, one can think of γ_k, which is distributed on $[L_k,B_k]$, as tax-inclusive costs, and γ_k^o, which is distributed on $[L_k^o,B_k^o]$, as tax-exclusive costs.

With this reinterpretation of our original model, one can immediately derive the following surprising result:

Theorem 2

Let $S_k(\tau)$ *denote the subset of high quality types in* Γ *that choose to signal with the* k-th *nondiscriminating activity. If* $\tau > \tau'$, *then* $S_k(\tau') \subset S_k(\tau)$.

Proof. By definition, $\lambda_k = (\gamma_k^o + \tau)/(L_k^o + \tau)$. Consequently,

(12) $$\frac{d\lambda_k}{d\tau} = \frac{L_k^o - \gamma_k^o}{(L_k^o + \tau)^2} \leq 0 \quad ,$$

with strict inequality when $\gamma_k^o > L_k^o$. The desired result then follows from theorem 1. ■

Theorem 2 indicates that the use of any particular nondiscriminating signal becomes more widespread when that signal is subject to a higher rate of taxation. Intuitively, nondiscriminating signals are "noisy" because they are contaminated by activity preferences. When the government subjects an activity to a publicly observable tax, this "purifies" the signal by reducing the size of the noise as a fraction of the total cost.

We depict the effect of taxation on activity choice graphically in figures I and II. For the purposes of these figures, we assume that the sender chooses between two nondiscriminating activities, the first of which may be subject to taxation. In a world with no taxes (figure I), the sender chooses to signal through z_1 whenever

(13) $$\gamma_2 \geq \gamma_1^o \frac{L_2}{L_1^o} \quad .$$

In contrast, in a world with taxes (figure II), the sender chooses to signal through z_1 whenever

(14) $$\gamma_2 \geq \left(\frac{L_2}{L_1^o + \tau}\right)(\gamma_1^o + \tau) \quad .$$

Notice that the line defining both constraints passes through the point (L_1^o, L_2), but constraint

(14) has a smaller slope. This means, as shown in figure II, that the tax on activity 1 enlarges the region of the preference space in which activity 1 is chosen.

In addition to studying the effects of taxation on the choice of a nondiscriminating activity, it is also of interest to investigate the impact of taxation on the levels of the signaling variables chosen in equilibrium. Implicit differentiation of (9) with respect to τ reveals that

$$\text{(15)} \qquad \frac{dx^*}{d\tau} = \left(\frac{L_k - \gamma_k}{L_k^2}\right)\left(\frac{V_x(x^*,R_H,Q_L)}{V_{xx}(x^*,R_H,Q_H) - \lambda_k V_{xx}(x^*,R_H,Q_L)}\right) < 0$$

(where, in signing this term, we have made use of the second order condition). Thus, for an agent who signals with the k-th nondiscriminating activity, an increase in the rate of taxation on activity k reduces the extent to which the agent signals through the discriminating activity, x. In effect, the "purification" of the signal resulting from the imposition of the publicly observed tax on z_k induces the agent to substitute away from x. It is then immediate from equation (10) that, for such an individual, the total amount of utility dissipated through z_k rises with the tax rate on z_k. The effect of the tax rate on the equilibrium level of z_k is, however, ambiguous. This reflects the offsetting effects of two factors: the higher tax rate induces the sender using z_k to substitute from x to z_k; however, a higher rate of taxation also implies that the sender can create the same penalty for prospective imitators through a smaller level of z_k.

If taxation renders a potential money-burning signal more attractive by "purifying" its content, then one might well wonder whether an increase in the tax rate yields a Pareto improvement.[7] The following result establishes that this conjecture is, in fact, correct.

Theorem 3

An increase in the tax rate τ is Pareto improving in the following sense: it strictly increases payoffs for high quality senders who use or switch to the taxed activity; it leaves the payoffs of all other senders unaffected; under Assumption R, it leaves the payoffs of receivers unaffected; and it strictly increases revenue.

Remark. It follows immediately from this theorem that one could achieve a strict increase in the payoffs for all agents simply by distributing the additional revenue through lump sum transfers, divided equally among all agents.

Proof. High quality senders who are induced to switch to the taxed activity as a result of an increase in τ must be better off -- otherwise, they would not switch. For those who burn money through activity k both before and after the tax increase, note that

$$\begin{aligned} \frac{d}{d\tau}U(x^*,z^*,R_H,Q_H,\gamma) &= V_x(x^*,R_H,Q_H)\frac{dx^*}{d\tau} - \frac{d}{d\tau}\left(\gamma_k z_k^*\right) \\ &= [V_x(x^*,R_H,Q_H) - \lambda_k V_x(x^*,R_H,Q_L)]\frac{dx^*}{d\tau} - z_k^* L_k \frac{d\lambda_k}{d\tau} \\ &> 0 \quad , \end{aligned} \tag{16}$$

where we have made use of equations (9), (10), and the fact that $d\lambda_k/d\tau < 0$. High quality senders who do not use or switch to activity k are clearly unaffected by the tax increase, as are low quality types and (under Assumption R) receivers.

There is clearly an increase in revenue raised from senders who switch to the taxed activity as a result of the tax increase. For senders who already used the taxed activity, revenues can be expressed as follows (using equation (10)):

(17) $$\tau z_k^* = \left(\frac{\tau}{L_k^o + \tau}\right)[V(x^*,R_H,Q_L) - V(0,R_L,Q_L)] \, .$$

The term $\tau/(L_k^o + \tau)$ is increasing in τ. Expression (15) indicates that x^* is decreasing in τ. Thus, the term in square brackets is also increasing in τ. Since both terms are strictly positive, the entire expression increases in τ. ■

Theorem 3 contrasts with the usual result for signaling models with money burning, where the effects of taxation are usually completely neutral (see e.g. Bernheim [1991] or Bagwell and Bernheim [1996]). The usual neutrality result follows from the fact that, in the absence of private information about rates of dissipation, taxes substitute perfectly for other costs. In the current context, taxes enhance welfare by purifying an imperfect signal.[8]

The possibility of Pareto improving taxes in this model raises an important question: why is the private sector unable to organize institutions that would remedy the original market failure in an equivalent fashion? Our analysis is driven by the assumption that the private sector cannot generate a money burning activity for which the true rate of dissipation is public knowledge. One possible justification for this assumption is that arrangements between private parties are never completely observable.

Imagine, for example, that, in the absence of taxation, an entrepreneur senses a profit opportunity, and sets him/herself up as a "private tax collector" (henceforth PTC). The PTC encourages senders to purchase its "services," which consist of exactly nothing, for a positive price. The PTC also widely publicizes a list of its customers, along with "volume" and price. Finally, the PTC hires an accounting firm to audit its transactions, and to certify the veracity of the published list. In this way, the PTC offers a private mechanism for dissipating verifiable quantities of resources.

Unfortunately for private markets, however, the PTC's services may not be viewed as clean signals. Imagine the consequences of competition among PTCs. A PTC would not be able to lure customers by lowering published prices, since the published price of its "service" is irrelevant (only expenditures matter). However, a PTC would be able to attract customers by maintaining published prices and granting secret price concessions. Obviously, each PTC would have an incentive to police its published prices as visibly as possible to establish credibility.[9] However, secret price concessions between the PTC and its customers would be very difficult to police privately. For example, it would be hard for outsiders to detect artificially inflated prices on other transactions between a customer and a PTC affiliate. The mere possibility that some PTCs might pay secret kickbacks implies that the true rate of resource dissipation for a PTC customer would be, at least to some degree, private information. Thus, the use of a PTC's services might be taken as signaling the existence of side deals, rather than quality. As long as the government can credibly commit not to kick back taxes secretly on a quid-pro-quo basis, the payment of taxes may provide a more believable means of dissipating a fixed quantity of resources.

Throughout this section, we have assumed that only one of the nondiscriminating activities is subject to taxation. One can easily introduce taxes on the other activities as well. All of the preceding analysis would still follow, with the exception of the statement concerning revenues in Theorem 3. In principle, an increase in the tax rate on a particular activity could reduce revenues by inducing senders to shift from more highly taxed money burning signals.

Even so, a sufficiently large increase in the rate of taxation on every nondiscriminating signal would still yield a Pareto improvement.

III. An Application to Corporate Dividend Policy

In this section, we explore an application of the general theory developed in section II to corporate dividend policy. Part A provides relevant background on the theory of corporate dividends. Part B elaborates on the application of our theory to corporate payout policy. Part C examines empirical evidence to determine whether the distinctive predictions of the theory are consistent with actual experience.

A. The Theory of Corporate Dividends

Traditionally, the widespread use of dividends as a method of distributing cash to shareholders has been regarded as puzzling. 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. 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.

Applications of signaling 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 [1987]). Yet few authors have ventured explanations for the practice of signaling with dividends rather than repurchases. Notable exceptions include John and Williams [1985] and Bernheim [1991].[10] In effect, both papers argue that it may be efficient to use nondiscriminating signals, such as dividends, in situations where it is optimal to burn money.[11]

John and Williams do not consider the possibility that firms might signal profitability by burning money in other forms, such as through charitable contributions, the construction of conspicuously expensive facilities, or advertising. Bernheim [1991] mentions this possibility, and attempts to resolve the issue by introducing -- as an extension to his basic model -- additional dimensions of private information, much as we have done in section II. In the context of corporate dividend policy, he obtains the counterpart of our Theorem 1. Unfortunately, his analysis of equilibrium activity levels and tax policy is confined to his basic model, in which the choice among nondiscriminating signals is implicitly indeterminant. As we have seen in section II, implications for tax policy and characterizations of equilibrium activity levels change when one moves from the basic money burning model (in which rates of resource dissipation are public knowledge) to a model in which the choice of a nondiscriminating signal is fully determined.

B. Dividends as Optimal Nondiscriminatory Signals

To understand the application of the current analysis in this context, it is necessary to formulate an explicit model of corporate payout policy. Consider a signaling environment in which the sender is a firm, and the receiver is a representative potential investor. The firm's message consists of K+1 distinct activities that reduce retained earnings, including repurchases, dividends, charitable contributions, and so forth. We will define x as the total amount of resources devoted to all of these activities (henceforth "total payouts"), and we will use the vector z will to describe the allocation of x across activities. In particular, z will be a K-dimensional vector, where the k-th element of z describes the level of resources devoted to the k-th activity, and where the level of the K+1-th activity is determined as a residual. Thus, the firm sends the "message" m = (x,z). We will arbitrarily take repurchases to be the K+1-th activity, so that the message m = (x,(0,0,...)) implies that the firm spends the amount x repurchasing shares, and does not engage in any other earnings-reducing activity. We will also take dividends to be the first element of z, so that the message m = (x,(x,0,0,...)) indicates that the firm pays a dividend of x, and does not engage in any other earnings-reducing activity.

Next, let Q denote the true (or full-information) value of the firm when m = (0,0). Even when investors correctly infer Q, if m ≠ (0,0) they will value the firm at less than Q, since the message m is costly (see below). In particular, we assume that the actual value of the firm is given by

$$W(x,z,Q) \equiv Q - c(x,Q) - \sum_{k=1}^{K} \delta_k z_k \quad . \tag{18}$$

In this formulation, total payouts (x) reduce the full information value of the firm by c(x,Q) (where $c_x(x,Q) > 0$), reflecting considerations such as the increased likelihood of costly bankruptcy. Each of the first K payout activities also wastes some of the distributed resources, relative to repurchases. Thus, if z_k represents dividends, then δ_k might measure the effective tax rate on dividends, relative to the effective tax rate on repurchases. Likewise, if z_k represents charitable contributions, and if these contributions are not valued by shareholders, then $\delta_k = 1$.

Upon observing m, investors will make inferences about the likelihood that the firm is high quality. Let ρ denote the subjective probability assessment of the representative investor that $Q = Q_H$. We assume that investors are risk-neutral, and that the value of the stock, P, is determined by competitive bidding among the investors, so that

(19) $$P(\rho,m) = \rho W(x,z,Q_H) + (1 - \rho)W(x,z,Q_L) \ .$$

We will assume that the manager cares about the current price of the firm, P, the true value of the firm, W, and possibly about the activities z_k. In particular, the payoff to the manager is given by

(20) $$\alpha P(\rho,m) + W(x,z,Q) + \sum_{k=1}^{K} \mu_k z_k \ .$$

At first glance, it may not be obvious that this model fits into our framework. This is because the preceding discussion does not explicitly identify the receiver's response. Moreover, if we think of the receiver's response as the "bid" $P(\rho,m)$, then, contrary to the assumptions in section II, this response depends directly upon the message m. However, since the process that translates inferences into bids is purely mechanical, one can think of the receiver's response as the formation of the belief ρ; that is, $R(\rho) = \rho$. In that case, the manager's (sender's) utility takes the form of equation (1), where

(21) $$V(x,R,Q) \equiv Q - c(x,Q) + \alpha R[Q_H - c(x,Q_H)] + \alpha(1 - R)[Q_L - c(x,Q_L)]$$

and

(22) $$\gamma_k \equiv (1 + \alpha)\delta_k - \mu_k \ .$$

Note that V is decreasing in x and strictly increasing in R, as required. Moreover, to the extent there is some $\hat{x}$ such that $c_x(x,Q_L) > c_x(x,Q_H)$ iff $0 < x < \hat{x}$, we will also have $V_x(x,R,Q_L)$

$< V_x(x,R,Q_H)$ iff $0 < x < \hat{x}$, as assumed in section II.[12]

Thus, the analysis of section II is directly applicable to our model of corporate payout policy. It follows that some fraction of firms will strictly prefer to use dividends as a nondiscriminating signal (Theorem 1). As Bernheim [1991] argues, this fraction will be large if managers have relatively little private information concerning γ_1 (the marginal cost of shifting the form of distribution from repurchases to dividends).

Although managers probably acquire relatively little private information concerning γ_1 compared with other γ_k, it is doubtful that all of their information about γ_1 is public. In general, managers probably learn about γ_1 through private contacts with shareholders. Large institutional shareholders may express their payout preferences directly to management, and these preferences may be driven in part by non-tax factors that are difficult for other investors to observe. A university seeking to increase the liquidity of its endowment might, for example, attempt to influence the dividend policy of firms in which its holdings are substantial.

Provided that managers do possess private information concerning γ_1, our analysis implies that the fraction of firms with positive dividends should rise with the dividend tax rate (Theorem 2). Increases in the dividend tax rate would also be Pareto improving. To establish this result, one invokes the portions of Theorem 3 that do not depend upon Assumption R. One then adds the observation that receivers are also unaffected by the dividend tax, since competitive bidding always drives investor surplus to zero.[13]

C. Empirical Analysis

Although our conclusions concerning the welfare effects of dividend taxation are provocative, they are not directly testable. The implications of Theorem 2 are, however, equally striking, and distinguish our model from other theories of corporate payout policy. Consequently, it should be possible to test our theory by examining the effect of dividend taxation on the use of cash dividends.

Previous authors have studied the relation between dividend taxation and the aggregate level of dividends in Britain (e.g. Feldstein [1970], Poterba and Summers [1985]) and the United States (e.g. Poterba [1987]). Generally, an increase in the relative taxation of dividends has been found to reduce the level of dividends in the aggregate. It is important to realize, however, that the aggregate change in dividends reflect a combination of two factors: a change in the level of dividends for dividend-paying firms, and a change in the set of firms that choose to pay dividends. In most models of corporate payout policy, these two effects are reinforcing. In our model, they work in opposite directions. While our model is certainly consistent with the observed negative relation between dividend tax rates and aggregate dividends, it has the striking implication that an increase in dividend taxation should nevertheless increase the fraction of firms paying dividends. In the remainder of this section, we examine empirical evidence to determine whether this distinctive prediction is consistent with actual experience.

Our empirical analysis is based on annual U.S. data for the period 1963-1988, inclusive. The dependent variable in our analysis, $FIRMTOT_t$, measures the fraction (normalized to lie between 0 and 1) of all companies incorporated within the U.S. with ordinary common shares listed on either the New York Stock Exchange (NYSE) or the American Stock Exchange (AMEX) that paid at least one cash dividend to shareholders during the year t. The required information is extracted from data tapes compiled by the Center for Research in Securities Prices (CRSP). CRSP differentiates cash dividends by frequency and tax status. For the bulk of our analysis, we use the broadest possible definition of cash dividends. However, we also explore the robustness of our results to alternative definitions, such as recurring (monthly, quarterly, semi-annual, or annual) "normal" payments to shareholders that are fully taxable as ordinary income to individuals.

The key independent variable in our analysis is the dividend tax penalty, θ_t, which is defined as follows:

$$\theta_t \equiv \sum_{j=1}^{s} w_{jt} \left(\frac{1 - m_{jt}}{1 - z_{jt}} \right) , \tag{23}$$

where m_{jt} is the marginal dividend tax rate at time t for investors in class j, z_{jt} is the accrual-equivalent capital gains tax rate at time t for investors in class j, w_{jt} are equity ownership weights, and s is the number of distinct shareholder classes.[14] We obtained historical data on θ_t from Poterba [1987].[15] Poterba's measures of equity ownership weights for households, pension funds, insurance companies, and banks are taken from the Federal Reserve Board flow of funds data. The distribution of dividend income across income classes within the household sector is based on Internal Revenue Service data. Poterba treats each income class as a separate shareholder category, and computes the marginal tax rate on dividend income for investors in each class. He constructs the capital gains tax rate by assuming that the effective accrual rate is approximately 0.25 times the statutory rate.

Additional explanatory variables are included to control for other economic conditions. As in essentially all previous studies of aggregate dividend levels, we include a measure of corporate profits, $PROF_t$. This variable expresses corporate profits as a percent of gross domestic product (the unit of measurement being percentage points). Our measure of corporate profits is obtained from the Citibase data files, and incorporates corrections for inflationary distortions of inventory valuation and capital consumption that result from historical cost-based accounting. In some specifications, we also control for the lagged value of this variable ($LPROF_t$).

It is noteworthy that voluntary dividend guidelines were adopted during the wage and price control period of the early 1970s. Poterba [1987] has found that these guidelines significantly depressed aggregated dividend levels. It is therefore important to allow, in some of our specifications, for the possibility that the guidelines also reduced the fraction of firms paying dividends. Thus, following Poterba [1987], we introduce an indicator variable, DG, which equals unity for the years 1972-74.[16]

Finally, we introduce several macroeconomic variables to control for possible business cycle effects on corporate dividend policy. These variables include capacity utilization expressed as a percentage of total capacity ($CAPUTIL_t$), the civilian unemployment rate ($UNEMP_t$), and the rate of growth in real gross domestic product computed from the fourth quarter of the preceding year to the fourth quarter of the current year ($GDPGRO_t$). The unit of measurement for each of these variables is percentage points. In certain specifications, we also examine the sensitivity of our results to the inclusion of a trend variable ($YEAR_t$).

Before turning to our findings, it is important to discuss our *a priori* expectations concerning the pattern of coefficients on variables other than θ_t. Since previous investigators have found a strong positive relation between dividends and corporate profits, one might expect a similar relation to exist between $FIRMTOT_t$ and $PROF_t$. However, if companies pay dividends to signal private information concerning profitability, there is little theoretical justification for this expectation.[17] This is because the variable $PROF_t$ is *public* information, whereas the fraction of firms paying dividends should be determined by the distribution of *private* information.

To understand this point, imagine a simple dividend signalling model in which there are only two types of firms ("low quality" and "high quality") and one method of distributing cash to shareholders (dividends). Let Π and $\Pi + \Delta$ (with $\Delta > 0$) denote, respectively, the levels of profits earned by low and high quality firms. Suppose that dividends are costly, and that the marginal cost of paying a dividend declines monotonically in retained earnings (so that the single crossing property is satisfied). Finally, let λ denote the fraction of firms that are high quality. For this model, $\Pi + \lambda\Delta$ is the analog of PROF. If PROF rises as the result of an increase in either Π, Δ, or λ, the average level of dividends increases (as observed in practice).[18] In contrast, FIRMTOT equals λ; changes in PROF arising from changes in Π or Δ have no effect on the fraction of firms paying dividends. Indeed, appropriately chosen changes in Π, Δ, and λ reduce the frequency of dividend usage while raising average profits.

When the single crossing property fails to hold and firms have many alternative methods of burning money (as in the more elaborate signalling model described in section III.B), the link between average profitability and the fraction of firms paying dividends becomes even more obscure. It difficult to see why the average level of profits across all firms would have an effect on the relative desirability of different nondiscriminatory methods of burning money (e.g. dividends vs. charity). Thus, our model (like other more standard signalling models) has no particular implication concerning the effect of $PROF_t$ on $FIRMTOT_t$.

Similar remarks apply with respect to our macroeconomic variables ($CAPUTIL_t$,

$UNEMP_t$, and $GDPGRO_t$). Generally, there is no reason to believe that more vigorous macroeconomic activity (as indicated by higher capacity utilization, lower unemployment, or higher GDP growth) would change the distribution of private information in a way that would systematically raise or lower the fraction of firms paying dividends. Indeed, even the theoretical link between macroeconomic activity and the *level* of dividends (for dividend-paying firms) is obscure: the marginal cost of paying dividends might rise during recessions because firms find themselves closer to insolvency, or it might rise during booms because the payment of a dividend requires firms to forego more profitable investment opportunities.

An inspection of the data reveals quite a strong positive correlation between the fraction of firms paying dividends, and the effective tax rate on dividends. The key series are depicted in figure III. During the first few years of our sample, the fraction of all listed firms paying dividends (FIRMTOT) appears to move in the same direction as θ. However, this pattern reverses in the late 60s. Between 1968 and 1979, θ falls and then rises, while FIRMTOT rises and then falls. Subsequent to 1979, θ rises and FIRMTOT falls monotonically. Overall, the correlation coefficient between FIRMTOT and θ is -0.5648. When only NYSE stocks are included, the correlation coefficient becomes -0.8547. The sign of this correlation is robust across subperiods. Since the 1980s witnessed several rounds of tax reform, a wave of mergers, and rapid growth of repurchase activity, one natural point to divide the sample is 1981/82. For the pre-1982 sample, the correlation coefficients are -0.2476 for all listed companies and -0.4534 for companies listed only on the NYSE; for the post-1981 sample, these correlation coefficients are -0.8721 and -0.8674, respectively.

When interpreting these results, it is important to keep in mind that θ declines when the dividend tax rate rises. Thus, the empirical patterns indicate that an increase in dividend taxation tends to be associated with an increase in the fraction of firms paying dividends. To the extent this finding proves robust when empirically tested while controlling for other relevant macroeconomic factors, it would be consistent with the implications of Theorem 2, and difficult to reconcile with other views of corporate dividend policy.

The results in table I measure the relation between dividend taxation and the fraction of companies paying dividends in the NYSE/AMEX sample, controlling for other explanatory variables.[19] The dependent variable (FIRMTOT) has been multiplied by 100 to scale the coefficients more conveniently. Each regression specification is initially estimated using ordinary least squares. However, the Durbin-Watson statistics generally indicate (with varying degrees of significance) positive autocorrelation of the OLS residuals in our sample. Therefore, each regression was re-estimated using the Cochrane-Orcutt iterative procedure to correct for AR(1) residuals. To verify that the Cochrane-Orcutt procedure is appropriate, the GLS residuals associated with the Cochrane-Orcutt estimates are then tested for first- and second- order autocorrelation using the Breusch-Godfrey LM test.

Equations (1) and (2) control for corporate profits, macroeconomic conditions, and the 1972-1974 dividend guidelines. The coefficients of θ are negative and significant at levels of confidence in excess of 99%, indicating that a higher dividend tax rate increases the fraction of companies paying dividends, precisely as predicted by our model. One can obtain some feel for the magnitude of this key coefficient through an example. Imagine that all investors face an effective marginal tax rate on capital gains of 0.20. Then an increase in the effective marginal tax rate on dividends (assumed for simplicity to be constant across investors) from 0.30 to 0.40

reduces θ from 0.875 to 0.750. Our estimates imply that a change of this magnitude would ordinarily be associated with a rise of 9.55 percentage points in the number of firms paying dividends.

The Breusch-Godfrey LM test was used as a diagnostic to determine whether the residuals continued to follow an AR(1) or AR(2) process after application of the Cochrane-Orcutt procedure. These statistics, which are distributed $\chi^2(1)$ for the AR(1) tests and $\chi^2(2)$ for the AR(2) tests (under the null hypothesis of no remaining autocorrelation) are presented, together with the associated p-values. The tests show no evidence of first or second-order autocorrelation surviving the Cochrane-Orcutt procedure. This finding validates statistical inferences based on the estimated coefficients and standard errors.

The coefficients of the other independent variables merit some comment. The fraction of firms paying dividends is negatively related to aggregate corporate profits, but the estimated coefficient is statistically insignificant. While this may at first seem peculiar, it is (for reasons discussed earlier in this section) entirely consistent with the general hypothesis that firms pay dividends to signal profitability. The pattern of coefficients for our macroeconomic variables reveals no clear relation between $FIRMTOT_t$ and the phase of the business cycle; $FIRMTOT_t$ rises with capacity utilization, but also rises with unemployment and falls with GDP growth. For the reasons discussed earlier, the absence of a clear pattern should not be regarded as surprising. Finally, we find that the dividend guidelines significantly depressed the fraction of companies that pay dividends, in keeping with Poterba's [1987] finding on the level of aggregate dividends.

In the remainder of table I, we explore the sensitivity of our results to alternative specifications of the dynamic process for dividend adjustment. Equations (3) and (4) follow much of empirical literature on factors determining the aggregate level of dividends by supplementing our basic specification with a measure of lagged profits. In principle, this allows for the possibility that dividends adjust gradually to changes in corporate profits. The estimated coefficients of LPROF are positive but insignificant, and the other coefficients (most notably that of θ) are little changed. The LM tests again show no evidence of a first-order autoregressive error process. While the test for a second-order autoregressive process is somewhat more suggestive of difficulties, the test statistic remains insignificant even at the 10% level. Equations (5) and (6) incorporate the YEAR variable to account for the possible existence of a time trend. The trend variable is insignificant, the estimated coefficients for θ_t increase slightly,[20] and there is again no evidence of a problematic error structure.

Table II contains additional results that further establish the robustness of our central findings. For equations (1) and (2), we reconstruct our dependent variable based on a more narrow definition of dividends -- regularly recurring (monthly, quarterly, semi-annual, or annual) "normal" payments to shareholders that are fully taxable as ordinary income to individuals. The coefficients of θ_t are estimated somewhat less precisely, but their magnitudes are little changed. When the Cochrane-Orcutt procedure is used (equation (2)), the coefficient of $PROFITS_t$ changes sign, but is still statistically insignificant.

For equations (3) and (4), we return to our original definition of dividends, but restrict attention to firms listed on the NYSE. There is only a negligible effect on the estimated coefficients (and associated standard errors) of θ_t. Once again, the coefficient on $PROFITS_t$ changes sign, but remains statistically insignificant.

Finally, for equations (5) and (6), we restrict attention to the set of firms that were traded continuously on either the NYSE or AMEX throughout the sample period. We undertake this exercise to explore the possibility that our results might be attributable to spurious changes in the composition of firms listed on these exchanges. This restriction significantly reduces the number of firms used to calculate $FIRMTOT_t$; while each data point in the original series was calculated using a sample of at least 2,000 firms, our fixed-firm series is based on a sample of 467 firms. The substantial reduction in sample size raises the possibility that the restricted sample may not be representative. Indeed, since the firms in this subsample tend to be older and more stable, they are significantly more likely to pay dividends in any given year. Nevertheless, results based on this sample provide further insight into the robustness of our key findings. Although the coefficients of θ_t are much smaller than in our basic specification, precision is also remarkably improved, and the estimated effects remain statistically significant at high levels of confidence. The smaller quantitative effect of tax policy could be attributable to the fact that older, more stable firms are less inclined to change their established dividend policies; nevertheless, the central result comes through. The other coefficients closely resemble those obtained in our basic specification, and there is again no evidence that the error structure is problematic.

We have performed a variety of other robustness exercises that, for the sake of brevity, are not reported in either table. For example, we have estimated separate equations using various combinations of the independent variables. Generally, our central findings emerge intact. For example, the estimated coefficient of θ_t is little affected by the omission of either the macroeconomic controls or the dividend guideline dummy variable, and its t-statistic continues to indicate a high level of statistical significance. However, in both cases there is evidence of residual first and second-order autocorrelation even after the application of the Cochrane-Orcutt procedure. Given the time series properties of the omitted variables, this is not surprising. Despite the fact that statistical inference is problematic for the unreported specifications, the robustness of the key coefficient provides us with considerable comfort.

IV. Conclusions

In this paper, we have proposed a theory of signaling via "money burning" that resolves each sender's indifference among alternative nondiscriminating signals. We establish three central results: (1) all potential money burning activities are used by some senders in equilibrium, (2) the taxation of a nondiscriminating signal is Pareto improving, and (3) an increase in the tax rate causes the use of the taxed activity to become more widespread. We have also applied this theory to the use of cash dividends as a signal of financial strength. The central testable implication of the model -- that an increase in the dividend tax rate should increase the number of firms that signal profitability by paying dividends -- has proven to be consistent with historical experience. This finding raises the possibility that the dividend tax may be Pareto improving.

Stanford University and Fordham University

References

Ambarish, Ramasastry, Kose John, and Joseph Williams, "Efficient Signaling with Dividends and Investments," *Journal of Finance*, XLII (June 1987), 321-343.

Bagwell, Laurie S., and B. Douglas Bernheim, "Veblen Effects in a Theory of Conspicuous Consumption," *American Economic Review*, (1996), forthcoming.

Bernheim, B. Douglas, "Tax Policy and the Dividend Puzzle," *Rand Journal of Economics*, XXII (1991), 455-476.

Bhattacharya, Sudipto, "Imperfect Information, Dividend Policy, and the 'Bird in the Hand' Fallacy," *Bell Journal of Economics*, X (Spring 1979), 259-270.

______, "Nondissipative Signaling Structures and Dividend Policy," *Quarterly Journal of Economics*, VC (August 1980), 1-24.

Cho, In-Koo, and David Kreps, "Signalling Games and Stable Equilibria," *Quarterly Journal of Economics*, CII (1987), 179-221.

Feldstein, Martin, "Corporate Taxation and Dividend Behaviour," *Review of Economic Studies*, XXXVII (1970), 57-72.

Fudenberg, Drew, and Jean Tirole. *Game Theory* (Cambridge, MA: MIT Press, 1991).

Hakansson, Nils H., "To Pay or Not to Pay Dividend," *Journal of Finance*, XXXVII (May 1982), 415-428.

Jensen, Michael C. and William H. Meckling, "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structures," *Journal of Financial Economics*, III (1976), 305-60.

John, Kose and David C. Nachman, "Multiperiod Financing Strategies and Dividend Management," CRSP Working Paper No. 210, 1987.

John, Kose and Joseph Williams, "Dividends, Dilution, and Taxes: A Signaling Equilibrium," *Journal of Finance*, XL (September 1985), 1053-1070.

Kumar, Praveen, "Shareholder-Manager Conflict, Costly Monitoring and the Information Content of Dividends," mimeo, Carnegie Mellon University, 1988.

Kumar, Praveen and Chester S. Spatt, "Asymmetric Information and the Smoothing of Dividends," mimeo, Carnegie Mellon University, 1987.

Milgrom, Paul, and John Roberts, "Price and Advertising Signals of Product Quality," *Journal* *of Political Economy*, XCIV (1986), 227-284.

Miller, Merton and Kevin Rock, "Dividend Policy under Asymmetric Information," *Journal of Finance*, XL (September 1985), 1031-1051.

Poterba, James M, "Tax Policy and Corporate Savings," *Brookings Papers on Economic Activity*, 2 (1987), 455-503.

Poterba, James M. and Lawrence H. Summers, "The Economic Effects of Dividend Taxation," in *Recent Advances in Corporate Finance*, E.I. Altman and M.G. Subrahmanyam, eds. (Homewood, IL: Richard D. Irwin, 1985), 227-284.

Rotemberg, Julio, "Pareto Improving Distortionary Taxes in the Presence of Signaling," Alfred P. Sloan School of Management Working Paper No. 2039-88, 1988.

Spence, A. Michael. *Market Signalling* (Cambridge, MA: Harvard University Press, 1974).

Williams, Joseph, "Efficient Signalling with Dividends, Investment, and Stock Repurchases," *Journal of Finance*, XLIII (1988), 737-747.

Endnotes

1. Milgrom and Roberts note this themselves [pp. 799-800], as do Bagwell and Bernheim [p. 28].

2. Separation along the dimensions of z is inconsequential, since the receiver does not care about γ.

3. Naturally, the model can also give rise to a variety of pooling equilibria. In many standard settings, no pooling equilibria survive application of the "equilibrium dominance" criterion proposed by Cho and Kreps [1987], which we use later to refine the set of separating equilibria. However, in this context, there may exist certain kinds of pooling equilibria that satisfy this and other refinements. We focus on separating equilibria because we have been unable to identify a class of pooling equilibria that is robust against refinements and reasonably simple in structure, for which existence is generally guaranteed. Although we leave the properties of pooling equilibria as open issues for further research, we conjecture that the key elements of our analysis will survive. In particular, an increase in the rate of taxation applied to a money burning activity should induce more agents to engage in the activity because it "purifies" the associated message (decreases the relative importance of private information).

4. In any separating equilibrium, low quality types are correctly identified as such. Consequently, they experience an equilibrium response of R_L. In a perfect Bayesian equilibrium, the receiver never chooses $R < R_L$ for any message m. If low quality types were assigned an equilibrium message involving positive levels of x or z, they could increase

their payoff by choosing $m = 0$. Consequently, they must choose $m = 0$.

5. For a fully separating equilibrium to exist, it must be the case that the maximized value of the objective function exceeds $V(0,R_L,Q_H)$. We will assume this throughout.

6. The second order condition requires that $V_{xx}(x,R_H,Q_H) - \lambda_k V_{xx}(x,R_H,Q_L) < 0$. This condition will be useful in the subsequent analysis.

7. In considering whether an increase in the tax rate yields a Pareto improvement, we abstract from the possibility that parties outside the model might benefit from the untaxed non-discriminating activities. In certain contexts, reductions in non-taxed discriminating activities (such as charity) might be harmful to third parties.

8. Theorem 3 bears some relation to the analysis of Rotemberg [1988], who demonstrated that a tax on a signal may be Pareto improving if the single crossing property fails to hold, and if senders are structurally prevented from burning money. We have demonstrated that the taxation of a signal may also be Pareto improving even when senders have access to money burning technologies, as long as there is some private information (unrelated to the characteristics being signaled) about true rates of resource dissipation.

9. Indeed, the manufacturers of conspicuous consumption goods often go to great lengths to assure that discounters do not degrade the signaling value of their products; see Bagwell and Bernheim [1994].

10. The analysis of John and Williams has been extended by Ambarish, John, and Williams [1987] and Williams [1988].

11. Despite this commonality, the models considered in these two papers differ in a number of other important respects. The relationships between them are discussed in Bernheim [1991].

12. As noted in Bernheim [1991], this property follows naturally from the presence of bankruptcy constraints.

13. It is natural to embellish the model by imagining a special class of investors who are the initial owners of the firms. In that case, initial owners of high quality, dividend paying firms will benefit from an increase in dividend taxation, while other initial owners will be unaffected.

14. By using data on θ_t that are constructed in this way, we implicitly assume that dividends are received in proportion to equity ownership. It is well known that floor traders, who are taxed equally on ordinary income and capital gains and losses, engage in tax arbitrage around ex days, and indeed some evidence indicates that these individuals receive a disproportionate share of dividends. Although it is therefore likely that θ_t overstates effective dividend taxation, movements in θ_t should nevertheless capture relative movements in the true effective dividend tax rate, particularly around major shifts in tax policy.

15. For 1987 and 1988 data on θ_t were obtained directly from James Poterba (private communication).

16. According to Poterba [1987], voluntary dividend controls were in effect between November 14, 1971, and April 30, 1974. The guidelines suggested that dividends should be limited to 4 percent above the highest payout level in the three years before the controls.

17. The expectation might be more reasonable under Jensen and Meckling's [1976] "free cash flow" hypothesis.

18. If Π rises, the marginal cost of dividends falls, and high quality firms must pay higher dividends to deter imitation. If Δ rises, imitation results in a larger share price increase, and high quality firms must therefore pay higher dividends to differentiate themselves. If λ rises, high quality firms pay the same level of dividends, but there are more high quality firms.

19. In these tables, we have adopted relatively simple functional specifications. As a practical matter, the relative brevity of our sample period dictates parsimony.

20. While the standard errors of these coefficients also rise, they remain statistically significant at high levels of confidence.

Table I
Determinants of Dividend Usage, NYSE/Amex Sample

Variable	OLS (1)	C-O (2)	OLS (3)	C-O (4)	OLS (5)	C-O (6)
θ	-76.43	-77.98	-78.36	-70.77	-85.65	-79.67
	(9.41)	(9.26)	(9.38)	(10.73)	(19.78)	(18.29)
PROFITS	-1.01	-0.44	-1.23	-1.01	-0.31	-0.34
	(0.84)	(0.82)	(1.18)	(1.20)	(1.57)	(1.43)
DG	-8.62	-8.06	-8.39	-7.56	-7.90	-7.97
	(1.36)	(1.36)	(1.25)	(1.52)	(1.92)	(1.85)
CAPUTIL	1.78	1.65	1.68	1.51	1.65	1.64
	(0.20)	(0.19)	(0.20)	(0.21)	(0.31)	(0.29)
UNEMP	4.74	4.68	4.85	4.84	4.48	4.64
	(0.53)	(0.53)	(0.51)	(0.63)	(0.72)	(0.70)
GDPGRO	-0.52	-0.52	-0.39	-0.34	-0.59	-0.53
	(0.27)	(0.23)	(0.31)	(0.29)	(0.30)	(0.26)
LPROFITS			0.74	0.41		
			(1.27)	(1.09)		
YEAR					0.13	0.02
					(0.24)	(0.22)
Constant	-43.83	-34.22	-37.25	-33.64	-37.81	-33.52
	(19.24)	(18.82)	(17.72)	(20.57)	(22.63)	(21.30)
F	34.44	32.04	36.51	20.15	28.45	26.47
Root MSE	1.8596	1.6356	1.6919	1.6577	1.8957	1.6826
DW	1.567		1.698		1.603	
ρ		0.1848		0.3167		0.1756
AR(1) test		0.5544		0.2001		0.5352
p-value		0.4565		0.6546		0.4644
AR(2) test		1.3294		3.7488		1.4559
p-value		0.5144		0.1534		0.4829

Standard errors in parentheses

Table II
Determinants of Dividend Usage, Examination of Robustness

Variable	OLS Narrow (1)	C-O Narrow (2)	OLS NYSE-only (3)	C-O NYSE-only (4)	OLS Fixed (5)	C-O Fixed (6)
θ	-78.05	-73.19	-75.75	-76.03	-22.66	-23.47
	(16.28)	(14.76)	(8.18)	(7.92)	(3.39)	(3.15)
PROFITS	-3.93	0.09	0.36	0.86	-1.47	-1.24
	(1.46)	(1.13)	(0.73)	(0.70)	(0.30)	(0.28)
DG	-11.25	-7.53	-3.38	-2.92	-3.92	-3.77
	(2.35)	(1.92)	(1.18)	(1.17)	(0.49)	(0.46)
CAPUTIL	2.10	1.22	0.83	0.75	0.78	0.74
	(0.34)	(0.28)	(0.17)	(0.16)	(0.07)	(0.06)
UNEMP	4.25	2.87	2.19	2.22	1.84	1.85
	(0.91)	(0.88)	(0.46)	(0.46)	(0.19)	(0.18)
GDPGRO	-0.28	-0.47	-0.47	-0.49	-0.12	-0.14
	(0.46)	(0.22)	(0.23)	(0.19)	(0.10)	(0.08)
Constant	-57.03	3.01	57.01	61.02	35.29	37.9
	(33.29)	(27.66)	(16.72)	(16.07)	(6.94)	(6.40)
F	12.38	9.81	30.64	28.72	40.25	42.86
Root MSE	3.2170	1.6880	1.6159	1.3669	0.6708	0.5729
DW	0.7844		1.325		1.603	
ρ		0.7577		0.2264		0.1276
AR(1) test		0.2064		0.1560		1.0320
p-value		0.6496		0.6929		0.3097
AR(2) test		4.0181		0.1817		2.2011
p-value		0.1341		0.9132		0.3327

Standard errors in parentheses

Figure I: Activity choice with no taxes

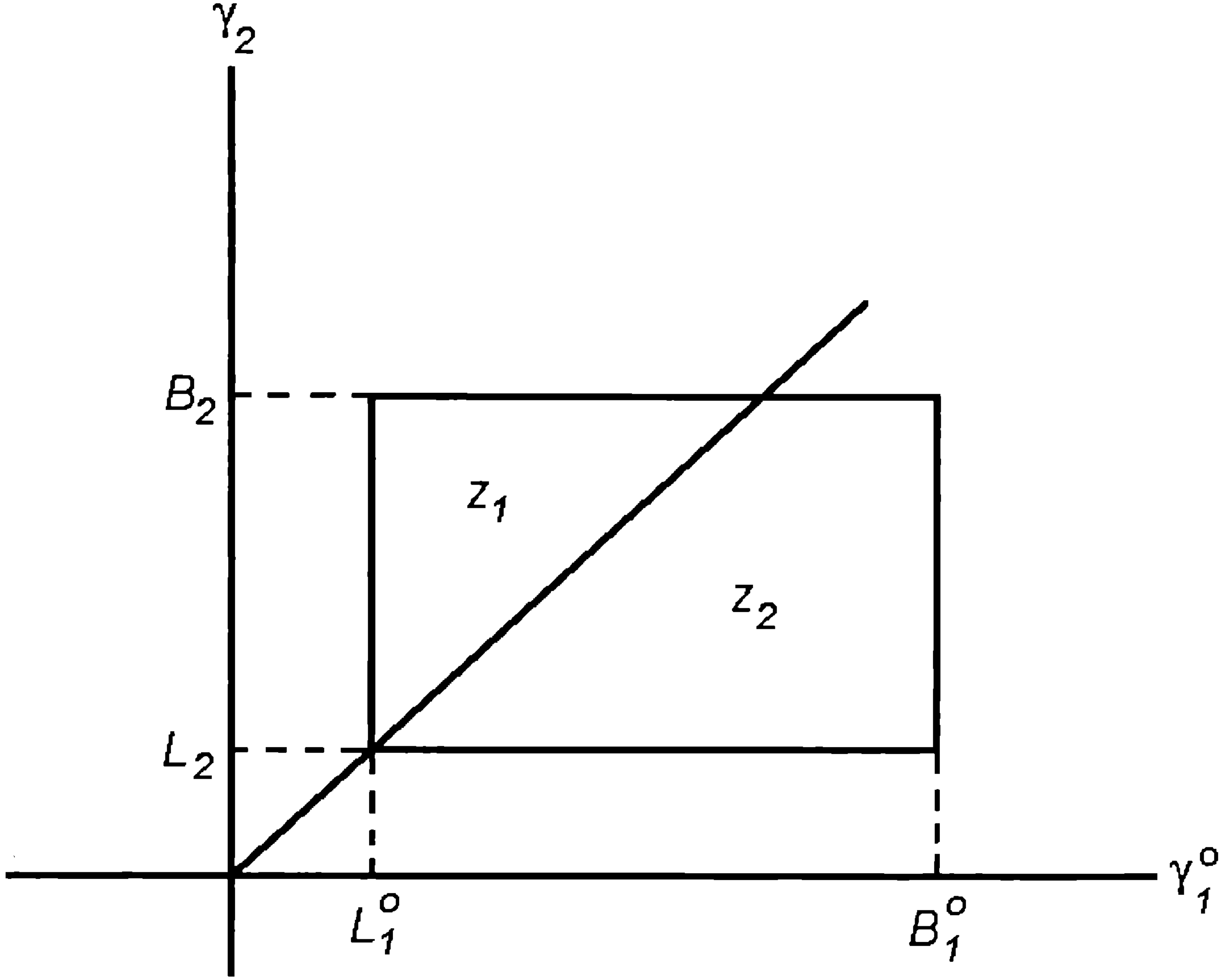

Figure II: Activity choice with a tax

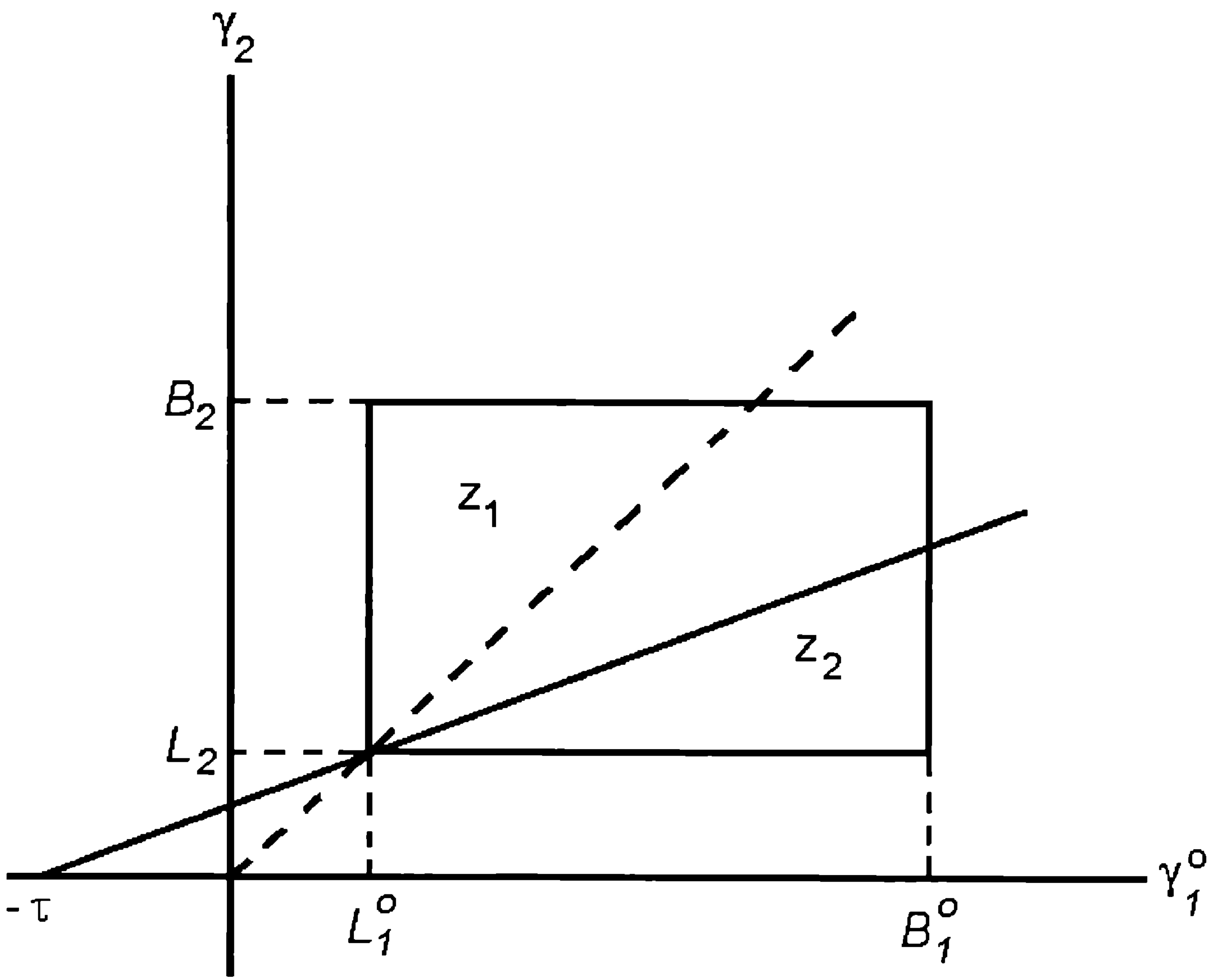

Figure III
Time Series of Key Variables