NBER WORKING PAPER SERIES

AN AGENCY THEORY OF DIVIDEND TAXATION

Raj Chetty
Emmanuel Saez

Working Paper 13538
http://www.nber.org/papers/w13538

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2007

We thank Alan Auerbach, Martin Feldstein, Roger Gordon, Kevin Hassett, James Poterba, and numerous seminar participants for very helpful comments. Joseph Rosenberg and Ity Shurtz provided outstanding research assistance. Financial support from National Science Foundation grants SES-0134946 and SES-0452605 is gratefully acknowledged. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

An Agency Theory of Dividend Taxation
Raj Chetty and Emmanuel Saez
NBER Working Paper No. 13538
October 2007
JEL No. G3,H20

ABSTRACT

Recent empirical studies of dividend taxation have found that: (1) dividend tax cuts cause large, immediate increases in dividend payouts, and (2) the increases are driven by firms with high levels of shareownership among top executives or the board of directors. These findings are inconsistent with existing "old view" and "new view" theories of dividend taxation. We propose a simple alternative theory of dividend taxation in which managers and shareholders have conflicting interests, and show that it can explain the evidence. Using this agency model, we develop an empirically implementable formula for the efficiency cost of dividend taxation. The key determinant of the efficiency cost is the nature of private contracting. If the contract between shareholders and the manager is second-best efficient, deadweight burden follows the standard Harberger formula and is second-order (small) despite the pre-existing distortion of over-investment by the manager. If the contract is second-best inefficient -- as is likely when firms are owned by diffuse shareholders because of incentives to free-ride when monitoring managers -- dividend taxation generates a first-order (large) efficiency cost. An illustrative calibration of the formula using empirical estimates from the 2003 dividend tax reform in the U.S. suggests that the efficiency cost of raising the dividend tax rate could be close to the amount of revenue raised.

Raj Chetty
Department of Economics
UC- Berkeley
521 Evans Hall #3880
Berkeley, CA 94720
and NBER
chetty@econ.berkeley.edu

Emmanuel Saez
University of California
549 Evans Hall #3880
Berkeley, CA 94720
and NBER
saez@econ.berkeley.edu

1 Introduction

The taxation of dividend income has generated substantial interest and controversy in both academic and policy circles. This paper aims to contribute to this debate by proposing a new theory of dividend taxation based on the agency theory of the firm (Jensen and Meckling 1976). Our model builds on two leading theories of dividend taxation and corporate behavior: the “old view” (Harberger 1962, 1966, Feldstein 1970, Poterba and Summers 1985) and the “new view” (Auerbach 1979, Bradford 1981, King 1977). The old view assumes that marginal investment is financed by the external capital market through new equity issues. Under this assumption, the taxation of dividends raises the cost of capital and, as a result, has a negative effect on corporate investment, dividend payouts, and overall economic efficiency. The new view assumes that marginal investment is financed from the firm’s retained earnings. In this case, the dividend tax rate does not affect the cost of capital because the dividend tax applies equally to current and future distributions. Therefore, the dividend tax rate does not affect the investment and dividend payout decisions of the firm, and has no effect on economic efficiency.[1]

There has been a longstanding debate in the empirical literature testing between the old and new views. Feldstein (1970), Poterba and Summers (1985), Hines (1996), and Poterba (2004) document a negative association between dividend payments and the dividend tax rate in the time series in the U.S. and U.K., consistent with the old view. In contrast, Auerbach and Hassett (2002) present evidence that retained earnings are the marginal source of investment funds for most corporations in the U.S., a finding that points in favor of the new view.

More recently, several papers have studied the effect of the large dividend tax cut enacted in 2003 in the U.S. (Chetty and Saez (2005), Brown et al. (2007), Nam et al. (2005)). Chetty and Saez documented four patterns: (1) Regular dividends rose sharply after the 2003 tax cut, with an implied net-of-tax elasticity of dividend payments of 0.75. (2) The response was very rapid – total dividend payouts rose by 20% within one year of enactment – and was stronger among firms with high levels of accumulated assets. (3) The response was much larger among firms where top executives owned a larger fraction of outstanding shares (see also Brown et al. (2007) and Nam et al. (2005)). (4) The response was much larger among firms with large

[1] See Auerbach (2003) for a summary of these models and the neoclassical literature on taxes and corporate behavior.

shareholders on the board of directors.

It is difficult to reconcile these four findings with the old view, new view, or other existing theories of dividend taxation. The increase in dividends appears to support the old view because dividends should not respond to permanent dividend tax changes under the new view.[2] However, the speed of the response is too large for a supply-side mechanism where dividend payouts rise because of increased investment eventually leading to higher profits and dividend payouts.[3] The rapid dividend payout response could be explained by building in a signaling value for dividends as in John and Williams (1985), Poterba and Summers (1985), or Bernheim (1991). However, neither the signaling model nor the standard old and new view models directly predict findings (3) and (4) on the cross-sectional heterogeneity in the dividend payout response by firm ownership structure.[4]

In this paper, we propose a simple alternative model of dividend taxation that matches the four empirical findings described above. The model is motivated by agency models of firm behavior that have been a cornerstone of the corporate finance literature since the pioneering work of Jensen and Meckling (1976), Grossman and Hart (1980), Easterbrook (1984), and Jensen (1986).[5] Our model nests the neoclassical old view and new view models but incorporates agency effects: managers have a preference for retaining earnings beyond the optimal level from the shareholders' perspective. We model this preference as arising from perks and pet projects, although the underlying source of the conflict between managers and shareholders does not matter for our analysis. Shareholders can provide incentives to managers

[2] One way of reconciling the dividend increase with the new view is if the tax cut was perceived as temporary by firms. However, Auerbach and Hassett (2005) document that the share prices of immature firms that are predicted to pay dividends in the future rose when the reform was announced, suggesting that firms perceived the tax cut as fairly permanent. In any case, the basic new view model would not explain findings (3) and (4) even for a temporary tax cut.

[3] Poterba's (2004) estimates using an old view model implied that the 2003 tax reform would increase dividend payments by 20 percent in the long run, but that the adjustment process would be slow, with only a quarter of the long-run effect taking place within three years.

[4] The empirical evidence is also not fully explained by Sinn's (1991) "life cycle" theory, which synthesizes the old view and new view in a model where firms start as old view firms and become new view firms once they have accumulated sufficient internal funds, at which point they start paying dividends. In that model, the payout response should be very small among mature firms with high levels of accumulated assets, but the data exhibit the opposite pattern.

[5] Several empirical studies have provided support for the agency theory as an explanation of why firms pay dividends (see e.g., Christie and Nanda 1994, LaPorta et al. 2000, Fenn and Liang 2001, Desai, Foley, and Hines 2007). Empirical studies have provided support for the predictions of the signalling theory of dividends as well (e.g. Bernheim and Wantz 1995, Bernheim and Redding 2001). See Allen and Michaely (2003) for a critical survey of these two literatures. It is of course possible that both signalling and agency effects are at play empirically.

to invest and pay out dividends through costly monitoring and through pay-for-performance (e.g. giving managers shares of the firm). Only the large shareholders of the firm choose to monitor the firm in equilibrium because of the free-rider problem in monitoring. Since managers have a higher preference for retained earnings than shareholders, they overinvest in wasteful projects and pay too few dividends relative to the first-best.

In this agency model, a dividend tax cut leads to an immediate increase in dividend payments because it increases the relative value of dividends for the manager and increases the amount of monitoring by large shareholders. Firms where managers place more weight on profit maximization – either because the manager owns a large number of shares or because there are more large shareholders who monitor the firm – are more likely to increase dividends both on the extensive and intensive margins in response to a tax cut. Hence, the agency model offers a simple explanation of the empirical findings from the 2003 tax cut and prior reforms that is consistent with evidence that marginal investment is funded primarily out of retained earnings.

The efficiency costs of dividend taxation in the agency model differ substantially from the predictions of neoclassical models. Since dividend taxation affects dividend payouts, dividend taxes always create an efficiency cost, irrespective of the marginal source of investment funds. The magnitude of the efficiency cost depends fundamentally on whether the contract between shareholders and managers is second-best efficient, i.e. if it maximizes total private surplus (excluding tax revenue) given the costs of monitoring and incentivizing the manager. If the contract is second best efficient, the efficiency cost of dividend taxation takes the standard Harberger triangle form, and is small (second-order) at low tax rates. An important implication of this result is that the pre-existing distortion of excessive investment by the manager does not *by itself* lead to a first-order deadweight burden from taxation. This result contrasts with the common view that the efficiency costs of taxing markets with pre-existing distortions are large (first-order). The conventional "Harberger trapezoid" intuition – which is based on a market with an exogenously fixed pre-existing distortion – breaks down when the size of the distortion is endogenously set at the second-best efficient level.[6]

However, if the contract between shareholders and the manager is *not* second best efficient,

[6] This result is related to the constrained first welfare theorem in economies with private information (Prescott and Townsend 1984a,b). If the first welfare theorem holds given the constraints, a small tax causes a second-order welfare loss.

dividend taxation does create a first-order efficiency cost. In our model, such second-best inefficiencies arise when companies are owned by diffuse shareholders. Each shareholder does not internalize the benefits of monitoring to other shareholders (a free-riding problem), and as a result monitoring is under-provided in equilibrium. In this situation, a corrective tax, such as a dividend subsidy, would improve efficiency. A dividend tax creates precisely the opposite incentive for the monitors, and leads to a first-order efficiency cost. Thus, when managers' interests differ from shareholders and companies are owned by diffuse shareholders – which is perhaps the most plausible description of modern corporations given available evidence (Shleifer and Vishny 1986, 1997) – dividend taxation can create substantial inefficiency.

Our analysis yields a simple yet fairly robust formula for the deadweight cost of taxation that nests the old and new view results. The formula is a function of a small set of empirically estimable parameters such as the elasticity of dividend payments with respect to the tax rate and the fraction of shares owned by executives and the board of directors. The formula is unaffected by allowing for equity issues, costly debt finance, or corruption within the board of directors.[7] We provide an illustrative calibration using estimates from the 2003 tax reform. The calibration shows that the marginal efficiency cost of raising the dividend tax from the current rate of 15% could be of the same order of magnitude as the amount of revenue raised. More than 80% of the efficiency cost arises from the agency effect rather than the Harberger channel emphasized in the old view model.

In addition to drawing heavily from well established models in the corporate finance literature, our analysis is related to two contemporaneous theoretical studies motivated by evidence from the 2003 dividend tax cut. Gordon and Dietz (2006) contrast the effects of dividend taxation in new view, signalling, and agency models. While our analysis shares some aspects with the model they develop, there are two important differences. First, since our framework is streamlined to focus exclusively on agency issues, we are able to derive additional results on firm behavior and efficiency costs.[8] Second, Gordon and Dietz assume that dividend payout decisions are always made by shareholders (as opposed to management) to maximize their total surplus. This assumption leads to very different results in both the positive and effi-

[7] An important limitation of the formula is that it does not account for share repurchases. Like most existing studies of dividend taxation in public finance, we abstract from share repurchase decisions. See section 5 for a discussion of how share repurchases would affect the efficiency results.

[8] We discuss the connections between our analysis and that of Gordon and Dietz in greater detail in sections 4 and 5.

ciency analysis. Gordon and Dietz's model does not directly predict a link between executive or board shareownership and behavioral responses to dividend taxation. In addition, taxing dividends does not create a first-order distortion in their model, since there is no pre-existing distortion and dividends are always set at the second-best efficient level. A second recent study is Korinek and Stiglitz (2006), who build on Sinn's (1991) model to analyze the effects of temporary changes in dividend tax rates. They incorporate financing constraints and establish new results on intertemporal tax arbitrage opportunities for firms. In contrast with our model, Korinek and Stiglitz assume that retained earnings are allocated efficiently by the manager. As a result, they obtain the new view neutrality result that permanent dividend tax policy changes have no effects on economic efficiency.

The remainder of the paper is organized as follows. In section 2, we present a simple two-period model that nests the old and new views as a benchmark reference. In section 3, we introduce agency issues into the model and characterize manager and shareholder behavior. In section 4, we characterize behavioral responses to dividend taxation in the agency model, and compare the model's predictions with available empirical evidence. In section 5, we analyze the efficiency consequences of dividend taxation in a set of agency models that make different assumptions about the formation of contracts between shareholders and managers. Section 6 provides an illustrative calibration of the general formula derived for deadweight burden. Section 7 concludes.

2 The Old and New Views in a Two-Period Model

We begin by developing a two-period model that nests the old view and new view, which serves as a point of departure for our agency analysis. Consider a firm that has initial cash holdings of X at the beginning of period 0. The firm can raise additional funds by issuing equity, which we denote by E. The firm's manager can do two things with the firm's cash holdings: pay out dividends or invest the money in a project that yields revenue in the next period. Let I denote the level of investment and $f(I)$ the revenue earned in period $t = 1$. Let $D = X + E - I$ denote the firm's dividend payment in period 0. In period 1, the firm closes and pays out $f(I)$ as a dividend to its shareholders. Assume that the production function f is strictly concave. A tax at rate t_d is levied on dividend payments in all periods. Investors can also purchase a government bond that pays a fixed interest rate of r (which is unaffected by the dividend tax rate), and therefore discount profits at a rate r.[9]

The manager's objective is to maximize the value of the firm, given by

$$V = (1-t_d)D + \frac{1-t_d}{1+r} f(X+E-D) - E \tag{1}$$

There are three choice variables: equity issues, dividend payments, and investment. To characterize how these variables are chosen, it is useful to distinguish between two cases: (1) A cash-rich firm, which has cash X such that $f'(X) \leq 1+r$ and (2) a cash-constrained firm, which has cash X such that $f'(X) > 1+r$.

Cash-Rich Firms – The New View. First observe that the firm will never set $E > 0$ and $D > 0$ simultaneously. If a firm both issued equity and paid dividends, it could strictly increase profits by reducing both E and D by \$1 and lowering its tax bill by \$$t_d$. Hence, any firm that wishes to raise additional funds will not pay dividends.

Now consider the marginal value of issuing equity when $D = 0$ for the cash-rich firm, which is given by

$$\frac{\partial V}{\partial E}(D=0) = \frac{1-t_d}{1+r} f'(X) - 1 \leq 0$$

Hence, a cash rich firm sets $E^* = 0$ and simply splits its prior cash holding X between investment and dividends: $I = X - D$. Now consider the optimal choice of dividends, denoted by D^*:

$$D^* = \arg\max_{D\geq 0} (1-t_d)\{D + \frac{f(X-D)}{1+r}\} \tag{2}$$

Hence the optimal dividend payout rate is determined by the first order condition

$$f'(X-D^*) = 1+r$$

Intuitively, firms invest to the point where the marginal product of investment $f'(I)$ equals the return on investment in the bond, $1+r$. We denote by I^S this socially efficient investment level. Note that the optimal dividend payment and investment level do not depend on the dividend tax rate t_d. This is the classic "new view" dividend tax neutrality result obtained by Auerbach (1979) and others. The source of this result is transparent in the two period case:

[9] Throughout this paper, we abstract from general-equilibrium effects through which changes in t_d may affect the equilibrium rate of return, r.

the $(1-t_d)$ term factors out of the value function in equation (1) when $E=0$. Intuitively, the firm must pay the dividend tax regardless of whether it pays out money in the current period or next period. As a result, dividend taxation has no impact on firm behavior and economic efficiency when firms finance the marginal dollar of investment out of retained earnings.

Cash-Constrained Firms – The Old View: Now consider a firm with X such that $f'(X) > 1+r$. The marginal value of paying dividends when $E=0$ for this firm is

$$\frac{\partial V}{\partial D}(E=0) = 1 - t_d - \frac{1-t_d}{1+r} f'(X) < 0.$$

Hence a cash-constrained firm never pays dividends in the first period: since the marginal product of investment exceeds the interest rate, it is strictly preferable to invest all retained earnings. This firm therefore invests all the cash it has: $I = X + E$. Now consider the optimal choice of equity issues, denoted by E^*:

$$E^* = \arg\max_{E \geq 0} \frac{1-t_d}{1+r} f(X+E) - E.$$

The optimal equity issue is given by the conditions

$$E^* = 0 \text{ if } (1-t_d)f'(X) < 1+r \tag{3}$$

$$(1-t_d)f'(X+E^*) = 1+r \text{ if } (1-t_d)f'(X) \geq 1+r \tag{4}$$

These conditions show that firms which finance their marginal dollar of investment from new equity issues invest to the point where the marginal *net-of-tax* return to investment $I^* = X + E^*$, equals the return on investment in the bond, $1+r$. Firms that have X sufficiently large so that $(1-t_d)f'(X) < 1+r$ have a net-of-tax return below the interest rate for the first dollar of equity. They therefore choose the corner solution of no equity (and no dividends, since they have $f'(X) > 1+r$) because of the tax wedge.

Implicit differentiation of equation (4) shows that $\partial I^*/\partial t_d < 0$ and $\partial E^*/\partial t_d < 0$ for firms at an interior optimum. For a sufficiently large tax cut, firms who were at the corner solution $E^* = 0$ begin to issue equity. These are the standard "old view" results that an increase in the dividend tax rate reduces equity issues and investment. The source of these results is again transparent in our simple two-period model: the $(1-t_d)$ term does *not* factor out of the value function in equation (1) when $D=0$ and $E>0$. When the marginal dollar of investment is financed from external funds, the price of a marginal dollar of investment is \$1 but the marginal product remains $(1-t_d)f'(I)/(1+r)$. A dividend tax increase therefore lowers the marginal product of investment but does not affect the price of investment for cash-constrained firms. Hence, an increase in t_d lowers the optimal amount of investment. This leads to lower dividends in the next period because revenue $f(I)$ falls. It is important to note, however, that dividend payments are *not* affected in the short-run in this simple old view model. Following a tax cut, investment increases immediately, and dividends increase only in the long-run after the additional investment pays off.

To calculate the efficiency cost of dividend taxation for cash constrained firms, denote by $P = D + f(I)/(1+r)$ total payout over the two periods. Total surplus in the economy is $W = V + t_d P$. The marginal deadweight burden of taxation is dW/dt_d. The envelope theorem applied to (1) implies that $dV/dt_d = -P$. Intuitively, since the firm has already maximized social surplus net of tax revenue, the only first-order effect of the tax on V is the mechanical revenue cost. This leads to the standard Harberger formula for deadweight burden:

$$dW/dt_d = -P + P + t_d \frac{dP}{dt_d} = -\frac{t_d}{1-t_d} \cdot \varepsilon_P \cdot P \tag{5}$$

where $\varepsilon_P = \frac{1-t_d}{P}\frac{dP}{d(1-t_d)}$ denotes the elasticity of total payout P with respect to the net-of-tax rate $1-t_d$.

Note that (5) characterizes deadweight burden for both cash-rich and cash-constrained firms. For cash-constrained (old view) firms, $P = f(I)/(1+r)$ falls with t_d and hence $\varepsilon_P > 0 \Rightarrow dW/dt_d < 0$. For new view firms, P does not respond to t_d and hence $\varepsilon_P = 0 \Rightarrow dW/dt_d = 0$.

Summary. These results are summarized on the left side of Table 1. In the neoclassical model of profit-maximizing firms, the efficiency consequences of dividend taxation depend critically on the marginal source of investment funds. Since most investment in developed economies is undertaken by firms with large amounts of retained earnings (Auerbach and Hassett 2002), the cash-rich case is perhaps most relevant in understanding the aggregate effects of dividend tax policy. This would imply that permanent changes in dividend tax policy have small effects on aggregate economic efficiency.

A key assumption underlying this conclusion is that firms' managers choose policies solely to maximize firm value. This assumption contrasts with the modern corporate finance literature, which emphasizes the tension between executives' and shareholders' interests in explaining corporate behavior. In the next section, we incorporate these agency issues into the model.

3 An Agency Model of Firm Behavior

In this and the next section, we restrict attention to a cash-rich firm that has $f'(X) > 1 + r$. Firms with $f'(X) < 1 + r$ never pay dividends. Since our goal is to construct a model consistent with available evidence on dividend payout behavior, it is the behavior of cash-rich firms that is of greatest interest for the positive analysis. In the efficiency analysis in section 5, we allow for cash-constrained firms, thereby nesting both the old and new view models when deriving formulas for deadweight burden. We defer modelling equity issues to section 5 since no cash-rich firm issues equity in equilibrium. To facilitate comparisons with the neoclassical model, the results for the agency model are summarized on the right side of Table 1.

3.1 Model Setup

The basic source of agency problems in the modern corporation is a divergence between the objectives of managers and shareholders. We model the source of the divergence as a "pet project" that generates no profits for shareholders but yields utility to the manager. In particular, the manager can now do three things with the firm's cash X: pay out dividends D, invest I in a "productive" project that yields profits $f(I)$ for shareholders, or invest J in a pet project that gives the manager private benefits of $g(J)$. Assume that both f and g are strictly concave.

The function g should be interpreted as a reduced-form means of capturing divergences between the managers' and shareholders' objectives. For example, the utility $g(J)$ may arise from allocation of funds to perks, tunnelling, a taste for empire building, or a preference for projects that lead to a "quiet life".[10] While there is debate in corporate finance about which of these elements of $g(J)$ is most important, the underlying structure that determines $g(J)$ does not matter for our analysis.

Manager's Objective. The agency problem arises because shareholders cannot observe

[10] There is a large literature in corporate finance providing evidence for such agency models. Recent examples include Rajan et al. (2000), Scharfstein and Stein (2000), and Bertrand and Mullainathan (2003).

real investment opportunities and hence have to let the manager decide about I, J, and D. Shareholders push managers toward profit maximization through two channels: incentive pay and monitoring. Incentive pay is achieved through features of the manager's compensation contract such as share grants, bonuses, etc. We model such financial incentives by assuming that the shareholders compensate the manager with a fraction α of the shares of the company.

Shareholders can also tilt managers' decisions toward profit maximization by monitoring. For example, shareholders could potentially veto some investment projects proposed by managers or pressure managers to pay dividends. More generally, monitoring can force managers to put more weight on the shareholders' objective to avoid being fired. To model the effect of such monitoring, suppose that γ units of monitoring makes the manager choose D, I, and J as if he values an extra dollar of profits by an extra \$$\gamma$ in addition to his direct share ownership. That is, the manager chooses I, J, and D to maximize

$$V^M = [\alpha(1-t_d)+\gamma]\cdot[D+\frac{1}{1+r}f(I)]+\frac{1}{1+r}g(X-I-D) \tag{6}$$

Let $\omega = \alpha(1-t_d)+\gamma$ denote the total weight that managers place on profits. When ω is low, the manager has little stake in the profits of the firm and is therefore tempted to retain excess earnings and invest in the pet project.[11]

Shareholders' Objectives. Next, we characterize how shareholders choose the level of monitoring (γ). Following Shleifer and Vishny (1986), we assume that the costs of monitoring are incurred by each shareholder who chooses to monitor the firm, whereas the benefits of better manager behavior accrue to all shareholders. This leads to a free-rider problem in monitoring. To model this problem, suppose that there are N shareholders, each of whom owns a fraction α_i of the shares (so that $\sum_1^N \alpha_i = 1-\alpha$). Each shareholder chooses a level of monitoring $\gamma_i \geq 0$. The total monitoring level is $\gamma = \sum \gamma_i$. Suppose that shareholders incur a fixed cost k if they choose to monitor the firm (i.e. set $\gamma_i > 0$). This fixed cost could reflect for example the cost of going to stockholder meetings. In addition, suppose there is a convex and increasing variable cost $c(\gamma_i)$ to do γ_i units of monitoring (the intensive margin) that satisfies $c'(\gamma_i = 0) = 0$. Increasing γ is costly because reviewing managers' plans, firing misbehaving

[11] The pet project $g(J)$ is presumably small relative to the firm's productive project $f(I)$. However, ω is also likely to be small in largely publicly traded corporations, where executives own a small fraction of total shares and diffuse shareownership can lead to a low level of monitoring. Combining a small pet project $g(J)$ with a small ω can make the manager deviate substantially from the shareholders' optimal investment level.

managers, etc. requires effort. Each shareholder chooses γ_i to maximize his net profits

$$V_i = (1 - t_d)\alpha_i \cdot [D + \frac{1}{1+r} f(I)] - k \cdot 1(\gamma_i > 0) - c(\gamma_i) \tag{7}$$

where $1(\gamma_i > 0)$ is an indicator function. In the Nash equilibrium, γ is determined such that each shareholder's choice of γ_i is a best response to the others' behavior. It is well known from the public goods literature that monitoring will be below the social optimum (i.e., the level that would be chosen if one shareholder owned the entire firm) in equilibrium.[12] In addition, it is easy to see that there is a threshold level $\overline{\alpha}$ such that small shareholders with $\alpha_i < \overline{\alpha}$ will not monitor the firm, while large shareholders with $\alpha_i > \overline{\alpha}$ do monitor. Since the number of large shareholders is typically small, it is natural to assume that these individuals cooperatively choose the level of monitoring γ by forming a "board of directors" that is in charge of monitoring the manager. Let α_B denote the total fraction of shares held by the board of directors. The board chooses γ to maximize its joint profits net of monitoring costs, recognizing that none of the small shareholders will ever participate in monitoring and taking into account the manager's behavioral responses:

$$V^B = (1 - t_d)\alpha_B \cdot [D(\omega) + \frac{1}{1+r} f(I(\omega))] - c(\gamma) \tag{8}$$

Ownership Structure. Thus far, we have specified the choices and objectives of the three key players in our agency model – the manager, the board of directors, and the small shareholders. What remains to be specified is the determination of the shares of these players – that is, how the firm's ownership structure (α and α_B) is set. We draw a distinction between the short-run positive analysis and the long-run efficiency analysis in the specification of the firm's ownership structure.

To understand the evolution of ownership structures, we use data on top executive share ownership from Execucomp and board of director share ownership from the Investor Responsibility Research Center for publicly traded firms in the U.S. See the data appendix for details on sample definition and construction of the share ownership variables. Figure 1 plots the logs of average managerial and (non-employee) board share ownership for the years in which data

[12] As emphasized in the corporate finance literature on free-riding problems, the Coasian solution (Coase 1960) is unlikely to emerge in this setting because of transaction costs in coordinating many small shareholders.

are available around the 2003 tax reform. For comparison, the log of total nominal dividend payments for firms listed in CRSP is plotted on the right scale. Note that the range of both scales is fixed at 0.8, facilitating direct comparisons. The figure shows a clear trend break in dividend payouts after the reform. Dividends rose by a total of 25% in the three years following the reform, after several years of remaining stable. In contrast, both ownership variables exhibit no trend break around the reform.

To quantify the effect of the tax reform on α, α_B, and D, we estimate a set of regression models. Since the variables plotted in Figure 1 all exhibit roughly linear time trends, we estimate models using OLS with two explanatory variables in Table 2: a linear year trend and a "post reform" indicator which is 1 for all years including and after 2003. The change in managerial or board shareownership following the reform is small and statistically insignificant, although the point estimates should be interpreted cautiously in view of the relatively large standard errors. In contrast, the post-reform dummy is large and statistically significant for total dividends. Consistent with the graphical evidence, these results suggest that the tax cut had little effect on ownership structure in the short run.

Since the evidence on dividend payout behavior we are attempting to explain concerns the effect of the 2003 dividend tax within a two year horizon, we take α and α_B as fixed in our positive analysis. In the longer run, and particularly when new firms are started, α and α_B are presumably endogenous to the tax regime. Therefore, in the efficiency analysis in section 5, we model how α and α_B are determined. Allowing for endogenous ownership structure is particularly important in the efficiency analysis because the deadweight cost of taxation depends critically on how α and α_B are determined.

3.2 Manager Behavior

Having set up the model, we now characterize manager and board behavior in the short run, taking ownership structure as fixed. The manager's behavior is determined by his weight on profits $\omega = \alpha(1 - t_d) + \gamma$. The manager chooses I and D to

$$\max_{I,D\geq 0} \omega[D + \frac{1}{1+r}f(I)] + \frac{1}{1+r}g(X - I - D). \tag{9}$$

Assume that $g'(0) > \omega f'(X)$, which guarantees an interior optimum in investment behavior. Then I and D are determined by the following first-order conditions:

$$\omega f'(I) = g'(X - I - D) \tag{10}$$

$$\omega \leq \frac{g'(X - I - D)}{1 + r} \text{ with strict equality iff } D > 0 \tag{11}$$

Let $D(\omega)$ and $I(\omega)$ denote the dividend and investment choices of the manager as a function of ω. To characterize the properties of these functions, define the threshold

$$\overline{\omega} = \frac{g'(X - I^S)}{1 + r} > 0,$$

Lemma 1 *$D(\omega)$ and $I(\omega)$ follow threshold rules:*

- *If $\omega \leq \overline{\omega}$ then $D(\omega) = 0$ and $I(\omega)$ is chosen such that $\omega f'(I) = g'(X - I)$.*
- *If $\omega > \overline{\omega}$ then $I(\omega) = I^S$ and $D(\omega) > 0$ is chosen such that $\omega = g'(X - I^S - D)/(1 + r)$.*

Proof. Consider $\omega \leq \overline{\omega}$. Suppose the firm sets $D > 0$. Then the first order conditions (11) and (10) imply that $f'(I) = 1+r$ and hence $I = I^S$. This implies $\omega = \frac{g'(X-I^S-D)}{1+r} > \frac{g'(X-I^S)}{1+r}$. It follows that $\omega > \frac{g'(X-I^S)}{1+r} = \overline{\omega}$, contradicting the supposition. Hence $\omega \leq \overline{\omega} \Rightarrow D(\omega) = 0$.

Now consider $\omega > \overline{\omega}$. Suppose the firm sets $D = 0$. Then the first order conditions (10) and (11) imply that $f'(I) \geq 1+r$ and hence $I \leq I^S$. This implies $\omega \leq \frac{g'(X-I)}{1+r} \leq \frac{g'(X-I^S)}{1+r}$. It follows that $\omega \leq \overline{\omega}$, contradicting the supposition. Hence $\omega > \overline{\omega} \Rightarrow D(\omega) > 0$, and (11) yields the desired expression for $D(\omega)$. QED.

Figure 2 illustrates the threshold rules that the manager follows by plotting $D(\omega)$, $I(\omega)$, and $J(\omega)$ with quadratic production functions. When ω is below the threshold value $\overline{\omega}$, the marginal value of the first dollar of dividends is negative in the manager's objective function. The optimal level of dividends is therefore zero, the corner solution. Intuitively, if managers have a sufficiently weak interest in profit maximization, they wish to retain as much money as possible for pet projects, and do not choose to pay out dividends. For ω above this threshold value, the managers choose a level of dividends that balances the marginal benefit of further investment in their pet project ($g'(X - I^S - D)/(1 + r)$) with the marginal benefit of paying out money and generating dividend income (ω). Above $\overline{\omega}$, increases in the weight on profits ω lead to increases in dividends and reductions in pet investment on the intensive margin:

$$D'(\omega) = -\frac{1+r}{g''(J(\omega))} > 0 \text{ for } \omega > \overline{\omega} \tag{12}$$

Now consider the manager's investment choice. When $\omega \leq \overline{\omega}$, the manager pays no dividends, and splits retained earnings between investment in the profit-generating project and the pet project. He chooses I to equate his private marginal returns of investing in the two projects, as in equation (10). An increase in ω increases productive investment I and reduces pet investment J:

$$I'(\omega) = -\frac{f'(I(\omega))}{\omega f''(I(\omega)) + g''(X - I(\omega))} > 0 \text{ for } \omega < \overline{\omega} \tag{13}$$

Once $\omega > \overline{\omega}$, the manager has enough cash to pay a dividend to shareholders. Since the marginal dollar of dividends could have been used for investment, he sets the investment level such that the marginal benefit of paying an extra dollar of dividends (ω) equals the marginal benefit of investing another dollar in the profit-generating project ($\omega f'(I)/(1+r)$). Hence the manager sets I such that $f'(I)/(1+r) = 1$, implying I is fixed at I^S for $\omega > \overline{\omega}$. Intuitively, the manager would only pay a dividend if his private return to further investment in the profitable project was below the interest rate. Since the tradeoff between dividends and profitable investment is the same for managers and shareholders, the manager only begins to pay a dividend once he has reached the optimal level of investment from the shareholder's perspective, I^S.

3.3 Board Behavior

In the short run, the board's only decision is to choose the level of monitoring. The board takes α_B as fixed and chooses γ to maximize

$$V^B = (1 - t_d)\alpha_B \cdot P(\omega) - c(\gamma) \tag{14}$$

where $P(\omega) = D(\omega) + f(I(\omega))/(1 + r)$ denote the firm's total payout as a function of ω. Because both D and P are (weakly) increasing in ω, $P(\omega)$ is also increasing in ω. The first order condition with respect to γ is:

$$c'(\gamma) = (1 - t_d)\alpha_B \cdot P'(\omega). \tag{15}$$

Intuitively, the board chooses γ such that the marginal increase in the board's share of profits by raising ω is offset by the marginal cost of monitoring. The second-order condition for an interior maximum is:

$$(1-t_d)\alpha_B \cdot P''(\omega) - c''(\gamma) < 0. \tag{16}$$

Since $c'(\gamma = 0) = 0$ by assumption, the optimal γ is always in the interior, and hence (16) must be satisfied at the optimal level of monitoring $\gamma(t_d)$.[13] This second-order condition turns out to be useful for the comparative statics analysis below.

4 Positive Analysis: Effects of Dividend Taxation

In this section, we analyze the effects of changes in dividend taxation on dividend payouts and investment behavior. Since the manager's behavior is fully determined by ω, for any variable $x \in \{D, I, J\}$,

$$\frac{dx}{dt_d} = \frac{dx}{d\omega}\frac{d\omega}{dt_d}$$

We have already characterized $\frac{dx}{d\omega}$ in the previous section. To characterize $\frac{d\omega}{dt_d}$, first observe that

$$\frac{d\omega}{dt_d} = -\alpha + \frac{d\gamma}{dt_d} \tag{17}$$

To calculate $\frac{d\gamma}{dt_d}$, implicitly differentiate the board's first-order-condition for γ in (15) to obtain:

$$\frac{d\gamma}{dt_d} = -\frac{\alpha_B[P'(\omega) + \alpha(1-t_d)P'']}{c'' - P'' \cdot \alpha_B(1-t_d)}. \tag{18}$$

Combining (17) and (18) leads to:

$$\frac{d\omega}{dt_d} = -\frac{\alpha_B P'(\omega) + \alpha \cdot c''}{c'' - P'' \cdot \alpha_B(1-t_d)} < 0. \tag{19}$$

The board's second-order condition for γ in (16) implies that the denominator of this expression is positive. The numerator is positive because P is increasing in ω and c is convex. Equation (19) therefore shows that a reduction in the dividend tax rate leads to an increase in the weight ω that managers put on profits. There are two channels through which this increase in ω occurs. First, a decrease in t_d mechanically increases the net stake $(1-t_d)\alpha$ that the

[13] The second order condition could hold with equality, a knife-edge case that we rule out by assumption.

manager has in the firm, effectively by reducing the government's stake (t_d) in the firm's profits. Second, a decrease in t_d generally increases the level of monitoring γ by the board.[14] Intuitively, monitoring rises because the return to monitoring is increased – since the external shareholders' net stake $(1-t_d)\alpha_B$ also rises when t_d falls – while the cost of monitoring is unchanged.

Given that $\frac{d\omega}{dt_d} < 0$, it is straightforward to characterize the short-run effect of dividend taxation on firm behavior. Since the manager follows a threshold rule in ω, changes in t_d lead to both intensive and extensive margin responses. We therefore analyze the effects of a discrete dividend tax cut from $t_d = t_1$ to $t_d = t_2 < t_1$ on a firm's behavior. Let $\Delta x = x(t_2) - x(t_1)$ denote the change in a variable x caused by the tax cut, and note that $\Delta\omega > 0$ from (19).

Proposition 1 *A dividend tax cut has the following effects on behavior for a cash-rich firm:*
(i) If $\omega(t_2) \leq \overline{\omega}$: $\Delta D = 0$, $\Delta I > 0$, $\Delta J < 0$, and $\Delta I + \Delta J = 0$.
(ii) If $\omega(t_1) < \overline{\omega} < \omega(t_2)$: $\Delta D > 0$, $\Delta I > 0$, $\Delta J < 0$, and $\Delta I + \Delta J < 0$.
(iii) If $\overline{\omega} \leq \omega(t_1)$: $\Delta D > 0$, $\Delta I = 0$, and $\Delta J < 0$.

Proof.

(i) When $\omega(t_2) \leq \overline{\omega}$, $D(t_2) = 0$ by Lemma 1. Since $\omega(t_2) > \omega(t_1)$, $D(t_1) = 0$ also. Therefore $\Delta D = 0$. Since $I + J + D = X$, and X is fixed, it follows that $\Delta I + \Delta J = 0$. Finally, it follows from (13) that $\frac{dI}{dt_d} = \frac{dI}{d\omega}\frac{d\omega}{dt_d} < 0$ when $\omega \leq \overline{\omega}$. Hence, $\Delta I > 0$ and $\Delta J = -\Delta I < 0$.

(ii) When $\omega(t_1) < \overline{\omega} < \omega(t_2)$, Lemma 1 implies $D(t_1) = 0$ while $D(t_2) > 0$. Hence $\Delta D > 0$. Since $\Delta D > 0$, $\Delta I + \Delta J = -\Delta D < 0$. By Lemma 1, $I(t_2) = I^S$ while $I(t_1)$ satisfies $\omega(t_1)f'(I(t_1)) = g'(X - I(t_1))$. Since $\omega(t_1) < \frac{g'(X-I(t_1))}{1+r}$ by (11), it follows that $f'(I(t_1)) > 1 + r = f'(I^S)$, which implies $I(t_1) < I(t_2)$. Hence $\Delta I > 0$ and $\Delta J = -\Delta D - \Delta I < 0$.

(iii) When $\overline{\omega} \leq \omega(t_1)$, $I(t_1) = I(t_2) = I^S$ because $\omega(t_2) > \omega(t_1)$. Equation (12) implies that $\frac{dD}{dt_d} = \frac{dD}{d\omega}\frac{d\omega}{dt_d} < 0$ when $\omega > \overline{\omega}$. Hence $t_2 < t_1 \Rightarrow \Delta D > 0$. Finally, $\Delta J = -\Delta D < 0$. QED.

Proposition 1 shows that the tax cut (weakly) increases dividend payments for all cash-rich firms because it raises the weight $\omega(t_d)$ that managers place on profits. The effect differs across

[14] Technically, it is possible to have $\frac{d\gamma}{dt_d} > 0$ if the third derivatives $g'''(J)$, $f'''(I)$, $c'''(\gamma)$ are sufficiently large in magnitude. When f, g, and c are quadratic, $\frac{d\gamma}{dt_d}$ is unambiguously negative. Hence, barring sharp changes in the local curvature of the production functions, monitoring falls with the dividend tax rate.

three regions of ω. For managers who place a very low weight on profits ($\omega(t_2) < \overline{\omega}$), paying any dividends is suboptimal after the tax cut, and hence before the tax cut as well. Hence, $\Delta D = 0$ for such firms. The second region consists of firms who were non-payers prior to the tax cut ($\omega(t_1) < \overline{\omega}$), but cross the threshold for paying when the tax rate is lowered to t_2. These firms initiate dividend payments after the tax cut. Finally, the third region consists of firms who had ω high enough that they were already paying dividends prior to the tax cut. The tax cut leads these firms to place greater weight on net-of-tax profits relative to the pet project, and therefore causes increases in the level of dividend payments on the intensive margin. Note that these changes in dividend payout policies occur in period 0 itself. This is consistent with the evidence that many firms announced dividend increases in the weeks after the 2003 tax reform was enacted (Chetty and Saez, 2005).

Now consider the effect of the tax cut on investment behavior. The tax cut increases the net-of-tax return to the profit-generating project while leaving the return to pet investment unaffected. As a result, the manager substitutes from investing in perks to the profit-generating project, and I (weakly) increases while J falls. In the first region, where $\omega(t_2) < \overline{\omega}$, total investment $(I + J)$ is unchanged, since $\Delta D = 0$ and total cash holdings are fixed. In the second region, where the firm initiates a dividend payment, investment in I rises to the socially efficient level I^S, while investment in J is reduced to finance the dividend payment and the increase in I. In this region, total investment falls when the tax rate is cut. Finally, when $\omega > \overline{\omega}(t_1)$, the manager maintains I at I^S and reduces investment in J to increase the dividend payment.

An interesting implication of these results is that a dividend tax cut weakly *lowers* total investment $I + J$ for cash-rich firms with an agency problem. Total investment, $I + J$, is the measure that is typically observed empirically since it is difficult to distinguish the components of investment in existing datasets. This prediction contrasts with the old view model, where a tax cut raises investment and with the new view model, where a tax cut has no effect on investment. Intuitively, a tax cut reduces the incentive for cash-rich firms to (inefficiently) over-invest in the pet project. It is important to note that the same result does not apply to cash-constrained firms in the agency model: A tax cut raises equity issues and productive (as well as unproductive) investment by such firms. Hence, a dividend tax cut leads to an (efficiency increasing) reallocation of capital and investment across firms, but its effect on aggregate investment is ambiguous. This result is potentially consistent with the large empirical literature on investment and the user cost of capital, which has failed to identify a robust relationship between tax rates and aggregate investment (see e.g., Chirinko 1993, Desai and Goolsbee, 2004).

Next, we examine how the effect of the tax cut on dividend payments varies across firms with different ownership structures. It is again useful to distinguish between extensive and intensive margin responses.

Proposition 2 *Heterogeneity of Dividend Response to Tax Cut by Ownership Structure:*
(i) Extensive Margin: Likelihood of Initiation. If $\omega(t_1) < \overline{\omega}$, initiation likelihood increases with α and α_B:

- *If $\Delta D > 0$ for α then $\Delta D > 0$ for $\alpha' > \alpha$*
- *If $\Delta D > 0$ for α_B then $\Delta D > 0$ for $\alpha'_B > \alpha_B$*

(ii) Extensive Margin: Size of Initiation. If $\omega(t_1) < \overline{\omega} < \omega(t_2)$: $\frac{\partial \Delta D}{\partial \alpha} > 0$, $\frac{\partial \Delta D}{\partial \alpha_B} > 0$.
(iii) Intensive Margin. If $\overline{\omega} \leq \omega(t_1)$ and g and c are quadratic: $\frac{\partial \Delta D}{\partial \alpha} > 0$, $\frac{\partial \Delta D}{\partial \alpha_B} > 0$.

Proof.

(i) The result follows directly from the effect of α and α_B on ω. Observe that

$$\frac{\partial \omega}{\partial \alpha} = (1 - t_d) + \frac{\partial \gamma}{\partial \alpha} = \frac{(1 - t_d)c''}{c'' - P'' \cdot \alpha_B(1 - t_d)} > 0.$$

using the second-order condition for γ in (16). Similarly,

$$\frac{\partial \omega}{\partial \alpha_B} = \frac{\partial \gamma}{\partial \alpha_B} = \frac{(1 - t_d)P'(\omega)}{c'' - P'' \cdot \alpha_B(1 - t_d)} > 0$$

Note that $\Delta D > 0$ at a given $\alpha \Rightarrow D(\omega(t_2, \alpha)) > 0$. Since $\frac{\partial \omega}{\partial \alpha} > 0$, we know that $\omega(t_2, \alpha') > \omega(t_2, \alpha)$. From (12), we have $\frac{\partial D}{\partial \omega} > 0$, which in turn implies $D(\omega(t_2, \alpha')) > D(\omega(t_2, \alpha)) > 0 \Rightarrow \Delta D > 0$ for α'. Exploiting the result that $\frac{\partial \omega}{\partial \alpha_B} > 0$ yields the analogous result for α_B.

(ii) When $\omega(t_1) < \overline{\omega} < \omega(t_2)$, $D(t_1) = 0$ and hence $\Delta D = D(t_2)$. It follows that $\frac{\partial \Delta D}{\partial x} = \frac{\partial D(t_2)}{\partial x} = \frac{\partial D}{\partial \omega}\frac{\partial \omega}{\partial x}$ for $x \in \{\alpha, \alpha_B\}$. We know that $\frac{\partial D}{\partial \omega} > 0$ from (12). Since $\frac{\partial \omega}{\partial \alpha} > 0$ and $\frac{\partial \omega}{\partial \alpha_B} > 0$ from (i), it follows that $\frac{\partial D(t_2)}{\partial \alpha} > 0$ and $\frac{\partial D(t_2)}{\partial \alpha_B} > 0$, which proves the claim.

(iii) When $\overline{\omega} < \omega(t_1)$, the dividend level is positive both at the initial and new tax rate and hence there is an intensive-margin response. Using equation (19), we have

$$\frac{\partial D}{\partial t_d} = \frac{dD}{d\omega} \cdot \frac{d\omega}{dt_d} = \frac{1 + r}{g''(J(\omega))} \cdot \frac{\alpha \cdot c'' + \alpha_B P'(\omega)}{c'' - P'' \cdot \alpha_B(1 - t_d)} \tag{20}$$

When $\overline{\omega} < \omega$, $P(\omega) = D(\omega) + \frac{f(I^S)}{1+r}$. Since $g''(J(\omega))$ is constant when g is quadratic and $D'(\omega) = P'(\omega) = -(1+r)/g''$, $P''(\omega) = D''(\omega) = 0$. Equation (20) therefore simplifies to

$$\frac{\partial D}{\partial t_d} = \frac{1+r}{g''}\left(\alpha - \alpha_B \frac{1+r}{g'' \cdot c''}\right)$$

Recognizing that c'' and g'' are constant and that $g'' < 0$, it follows that $\frac{\partial D}{\partial t_d}$ is constant in t_d and decreasing in α and α_B. Therefore, $\frac{\partial^2 D}{\partial t_d \partial \alpha} < 0$ and $\frac{\partial^2 D}{\partial t_d \partial \alpha_B} < 0$. Finally, $\Delta D = (t_2 - t_1)\frac{\partial D}{\partial t_d}$ with $t_2 - t_1 < 0$, and therefore $\frac{\partial \Delta D}{\partial \alpha} > 0$ and $\frac{\partial \Delta D}{\partial \alpha_B} > 0$. QED.

Figure 3a plots D against α in two tax regimes, with $t_1 = 40\%$ and $t_2 = 20\%$. The figure illustrates the three results in Proposition 2. First, among the set of firms who were non-payers prior to the tax cut, those with large executive shareholding (high α) are more likely to initiate dividend payments after the tax cut. This is because managers with higher α are closer to the threshold ($\overline{\omega}$) of paying dividends to begin with, and are therefore more likely to cross that threshold. Second, conditional on initiating, firms with higher α initiate larger dividends. Since $D(t_2)$, the optimal dividend conditional on paying, is rising in α, the size of the dividend increase, $\Delta D = D(t_2)$, is larger for firms with higher values of α in this region. Third, among the firms who were already paying dividends prior to the tax cut, the intensive-margin increase in the level of dividends is generally larger for firms with higher α.[15] Intuitively, the manager's incentives are more sensitive to the tax rate when he owns a larger fraction of the firm. Since a change in t_d has a greater effect on ω when α is large, the change in dividends is larger.

These three results apply analogously to the board's shareholding (α_B), as shown in Figure 3b. Non-paying firms with large α_B are closer to the threshold $\overline{\omega}$, and are thus more likely to initiate dividend payments following a tax cut. In addition, the board's incentives to monitor the firm are more sensitive to the tax rate when it owns a larger stake in the firm. A change in t_d has a greater effect on γ when α_B is large, leading to a larger dividend response.

All of these predictions regarding the impact of ownership structure on dividend payout responses are consistent with evidence from the 2003 tax cut. This is because in our agency model, managers choose the level of dividends and the board (rather than shareholders at large) sets monitoring. In contrast, Gordon and Dietz's (2006) agency model assumes that

[15] As above, this result holds as long as there are no sharp changes in the local curvature of the production functions. If $g'''(J)$ and $c'''(\gamma)$ are sufficiently large in magnitude, it is possible to have $\frac{\partial^2 D}{\partial t_d \partial \alpha_B} > 0$.

dividends are picked by the board, who represent the interest of all shareholders. Hence, their model does not directly explain the empirical finding that firms with large manager or board ownership were more likely to increase dividends following the tax cut. Their model does, however, generate the empirically validated prediction that dividends change slowly over time. In this sense, our model and Gordon and Dietz's analysis should be viewed as complementary efforts to explain different aspects of dividend policies.

Auxiliary Predictions. Our model predicts that firms with more assets and cash holdings (higher X) are more likely to initiate dividend payments following a tax cut.[16] In contrast, neoclassical models that nest the old and new views (e.g. Sinn 1991) predicts that firms with higher assets will respond *less* to a tax cut because they are more likely to finance marginal investment out of retained earnings. Chetty and Saez (2005) document that firms with higher assets or cash holdings were more likely to initiate dividends after the 2003 tax reform, consistent with the agency model.

The importance of the interests of "key players" (executives and large external shareholders) is underscored by Chetty and Saez's finding that firms with large *non-taxable* shareholders (such as pension funds) were much less likely to change dividend payout behavior in response to the 2003 tax reform. Although we have not allowed for heterogeneity in tax rates across shareholders in our stylized model, it is easy to see that the introduction of non-taxable shareholders would generate this prediction. If the board includes non-taxable large shareholders, a given change in t_d has a smaller impact on the board's incentive to increase monitoring. As a result, the tax cut causes a smaller increase in γ and generates smaller ΔD.[17]

5 Efficiency Cost of Dividend Taxation

In this section, we develop formulas for the deadweight burden of dividend taxation in the agency model. The efficiency consequences of taxation depend on how the firm's ownership structure (α and α_B) are determined. When both α and α_B are endogenous, it is convenient

[16] Firms with higher X are closer to the threshold of paying dividends, for two reasons: (1) $\overline{\omega}$ is falling in X and (2) γ is rising in X. A tax cut is therefore more likely to make firms with higher X cross the threshold and initiate dividend payments.

[17] By assuming that all shareholders are taxed equally at rate t_d, our model also ignores tax clientele effects. Allen, Bernardo, and Welch (2000) propose a theory of tax clienteles in which firms strategically pay dividends to attract large shareholders as monitors. It would be interesting to explore the effects of dividend tax changes in such a model in future work.

to write the formulas in terms of $\beta_B = \frac{\alpha_B}{1-\alpha}$ – the fraction of external shares held by the board – rather than α_B. For expositional simplicity, we consider three models of increasing generality. First, we consider the case where β_B is fixed at 1, i.e. the firm is owned by a single external shareholder ($\beta_B = 1$) who chooses the manager's share α. We then consider the case where β_B is fixed at a value less than one. Finally, we analyze a model where β_B is endogenously determined and firms can issue new equity. We present a general formula for excess burden that nests the three cases in the third subsection. In the appendix, we show that two further extensions – corruption of the board and debt finance – yield very similar formulas.

5.1 Single External Shareholder [$\beta_B = 1$]

Determination of α. We model the determination of managerial share ownership using the standard principal-agent framework in the corporate finance literature. The shareholder chooses the manager's stake α to maximize firm value, taking into account the manager's aversion to risk and his participation constraint. To model risk aversion, it is necessary to introduce uncertainty into the firm's payoff. Suppose that the profit-generating project now has a payoff $f(I)$ only with probability π; with probability $1-\pi$ it returns 0. The manager also receives a salary payment S independent from the profit outcome. The salary S is expensed to the firm, i.e. deducted from profits before dividend payments and dividend taxes are paid. We introduce this salary payment so that the shareholder can meet the manager's participation constraint as described below.

Let $u(c)$ denote the manager's consumption utility, which we assume is strictly concave. In addition to this consumption utility, the manager continues to get utility from the pet project g. With this notation, the manager's expected utility is given by

$$\mathrm{E}u(D,I) = \pi u\left(\alpha(1-t_d)[D + \frac{1}{1+r}f(I) - S] + S\right) + (1-\pi)u\left(\alpha(1-t_d)[D - S] + S\right) \\ + \frac{1}{1+r}g(X - I - D - S)$$

Let $\beta = \alpha(1-t_d)$ denote the manager's net of tax share of profits. As above, we assume that the manager's total weight on profits is augmented by shareholder monitoring (γ), so that the manager's objective is to maximize

$$\pi u\left((\beta+\gamma)[D+\frac{1}{1+r}f(I)-S]+S\right)+(1-\pi)u\left((\beta+\gamma)[D-S]+S\right)+\frac{1}{1+r}g(X-I-D-S)$$

The manager's maximization program generates a mapping from β, γ, and S to a choice of investment and dividends, which we denote by the functions $I(\beta, \gamma, S)$ and $D(\beta, \gamma, S)$. Note that t_d affects the manager's choices only indirectly through its effects on β, γ, and S.

The shareholder, who has linear utility, chooses β, γ, and S to maximize his net payoff

$$W_S = (1-t_d-\beta)[D(\beta,\gamma,S)+\frac{1}{1+r}\pi f(I(\beta,\gamma,S))-S]-c(\gamma) \tag{21}$$

subject to the participation constraint of the manager

$$\mathrm{E}u(D, I) = 0.$$

The participation constraint pins down the value of S given a choice of β and γ, so we can write $S = S(\beta, \gamma)$. Let total expected profits be denoted by

$$P(\beta,\gamma) = D(\beta,\gamma,S(\beta,\gamma)) + \pi f[I(\beta,\gamma,(\beta,\gamma))]/(1+r) - S(\beta,\gamma).$$

With this notation, the shareholder's problem reduces to choosing β and γ to maximize:

$$W_S = (1-t_d-\beta)P(\gamma,\beta)-c(\gamma) \tag{22}$$

If the manager's utility were linear, the shareholder would achieve his objective by setting $\alpha = 1$ and S sufficiently negative ("selling the firm to the manager"), so that incentives of the principal and agent are perfectly aligned. When the manager is risk averse, however, increasing pay-for-performance (α) while keeping expected the manager's utility constant forces the principal to raise total expected compensation. Since raising compensation reduces net profits, the optimal α is less than 1 when the manager is risk averse.

Efficiency Analysis. Since the manager's surplus is pinned at zero by his participation constraint, total surplus in the economy (W) is the sum of the shareholder's welfare and government revenue:

$$W = t_d P + W_S$$

To calculate the efficiency cost of taxation, observe that $dW_S/dt_d = -P$ because β and γ have already been optimized by the shareholder – a simple application of the envelope theorem. As a result,

$$\frac{dW}{dt_d} = P + t_d \frac{dP}{dt_d} - P = t_d \frac{dP}{dt_d} = -\frac{t_d}{1-t_d} P \cdot \varepsilon_{P,1-t_d} \tag{23}$$

where

$$\varepsilon_{P,1-t_d} = \frac{dP}{d(1-t_d)} \cdot \frac{1-t_d}{P}$$

is the elasticity of total dividend payouts with respect to the net-of-tax rate. This expression shows that the efficiency cost of dividend taxation is positive even for cash-rich firms in the presence of agency problems. The formula for deadweight burden coincides exactly with the Harberger formula obtained from the old view model. Deadweight burden is a second-order function of the tax rate. That is, the marginal deadweight cost of taxation is small at low tax rates.

First-Order vs. Second-Order Efficiency Costs. In view of the positive result that dividend taxation reduces dividend payouts, it is not surprising that dividend taxation has an efficiency cost in the agency model. The reduction in dividend payments reduces government revenue from dividends (a fiscal externality), and leads to an efficiency cost through the standard Harberger channel.

The more surprising aspect of equation (23) is the magnitude of the deadweight burden. A basic intuition in the theory of taxation is that taxing a market with a pre-existing distortion leads to a first-order efficiency cost (see e.g. Auerbach 1985, Hines 1999, Auerbach and Hines, 2003, Goulder and Williams 2003, Kaplow 2008). That is, introducing a small tax in a previously untaxed market leads to a large deadweight burden if the market for that good is already distorted. Equation (23) appears to violate this principle, since it predicts a second-order deadweight burden from dividend taxation despite the pre-existing distortion in the agency model. In particular, the manager under-provides dividends relative to the first-best social optimum, and taxing dividends further reduces dividend payments.

Why does our result differ from that of other studies in the tax literature? The pre-existing distortions analyzed in the studies cited above are exogenously fixed. In contrast, the pre-existing distortion in our model is endogenously determined through optimization of the manager's contract to maximize total private surplus given the informational constraints.

This endogenous determination of the distortion makes the deadweight cost of taxation second-order.

To understand the connection between deadweight burden and endogenous distortions, it is useful to analyze the problem at a more abstract level. Consider an economy in which the private sector has a vector x of choices. Total private surplus in the economy $W_S(x, t_d)$ is a function of the private choices x and the government tax rate t_d. For example, in the model we have just studied $x = (\beta, \gamma)$ and $W_S(x, t_d) = (1 - t_d - \beta)P(\gamma, \beta) - c(\gamma)$. It is important to note that the choice vector x can be parametrized in alternative but equivalent ways. We could have defined $x = (\alpha, \gamma)$ and written $W_S(x, t_d) = (1 - t_d)(1 - \alpha)P(\gamma, \alpha(1 - t_d)) - c(\gamma)$. The following lemma provides sufficient conditions for deadweight burden to have the Harberger form in such an economy.[18]

Lemma 2 *A tax has a second-order efficiency cost if both of the following conditions are satisfied.*
(i) [No intrinsic government advantage] There exists a parametrization of the choice vector x such that $\frac{\partial W_S}{\partial t_d}|_x = -P$ for all x and t_d.
(ii) [Second-best efficiency] Private agents choose x to maximize $W_S(x, t_d)$.

Proof. Total social surplus is $W = t_d P(t_d) + W_S(x(t_d), t_d)$. Hence

$$\frac{dW}{dt_d} = P + t_d \frac{dP}{dt_d} + D_x W_S \cdot \frac{dx}{dt_d} + \frac{\partial W_S}{\partial t_d}|_x. \tag{24}$$

Condition (ii) implies that $D_x W_S = 0$ and hence the third term in(24) is zero. Condition (i) implies that the fourth term is $-P$. Therefore, (24) simplifies to $dW/dt_d = t_d dP/dt_d$. This is the standard Harberger formula, implying that a small tax has a second-order efficiency cost. QED.

The first condition in Lemma 2 requires that the government and private market have the same tools to resolve informational constraints. Aside from the mechanical reduction in private welfare due to the tax increase, the effect of any change in t_d on W_S must be replicable by changes in private market contracts. The formal statement $\frac{\partial W_S}{\partial t_d}|_x = -P$ captures this intuition by requiring that with an appropriate specification of the private sector's

[18]Abstractly, this lemma can be viewed as a consequence of the constrained first welfare theorem in economies with private information (Prescott and Townsend 1984a, b).

tools, the government can do nothing to affect private welfare beyond mechanically extracting revenue.[19] The second condition is that the private market choices maximize total private surplus. Lemma 2 shows that as long as these two conditions hold, deadweight burden is second-order even in the presence of pre-existing distortions. Intuitively, since t_d and x affect W_S in the same way (net of revenue effects) and x has already been optimized, any change in t_d has only a second-order effect on social welfare.

In our model, condition (i) is satisfied with the parametrization $x = (\beta, \gamma)$. With this parametrization, it follows immediately that the tax t_d has no effect on the contracting possibilities between the shareholder and the manager, because the manager's decision rule can be expressed purely as a function of β and γ (and not t_d). In particular, shareholders can fully replicate the effect of t_d on the manager's behavior by varying α. Since the single shareholder chooses α to maximize total private welfare (condition ii), the only first-order effect of t_d on W_S is the mechanical effect of paying more taxes ($-P$). This first-order effect is exactly offset by the mechanical increase in tax revenue (P) in $W = t_d P + W_S$. This leaves only the term arising from the behavioral response of P to t_d, which reduces government revenue. This term is proportional to t_d, explaining why deadweight burden is second-order.

In contrast with our model, models with exogenous pre-existing distortions effectively assume that the private sector has no tool to affect the size of the distortion, violating condition (i). In that setting, the government has a technological advantage relative to the private sector, and can induce first-order changes in efficiency through taxes and subsidies. The same applies in our analysis: if α were exogenously fixed at some level α_0 not chosen to maximize the shareholder's welfare, the efficiency cost of dividend taxation has an added $(1 - \alpha_0)\frac{\partial P}{\partial t_d}$ term and is therefore first-order. Intuitively, when α is fixed, the government can change the manager's weight on profit maximization $\omega = \alpha(1 - t_d) + \gamma$ costlessly through changes in t_d, whereas the private sector must rely on the costly γ mechanism. This advantage for the government leads to a first-order efficiency gain from a dividend subsidy, and hence a first order efficiency cost from dividend taxation.

The general lesson – which is of relevance beyond dividend taxation – is that identifying a pre-existing distortion is *not* sufficient to infer that government taxes or subsidies will have

[19] When the private sector consists of a single maximizing agent (e.g. as in the old view model), there are no contracting issues, and this condition is trivially satisfied.

first-order effects on welfare, contrary to conventional wisdom. It is critical to understand the private sector's ability to affect the size of the distortion – specifically whether the private sector has the same tools as the government and whether the private sector reaches the second-best efficient outcome.

In the context of dividend taxation, there is no obvious reason that government taxation or subsidies are a superior method of resolving agency problems in firms.[20] Hence, all the models we will consider satisfy the first condition in Lemma 2. However, the second condition is likely to break down in an economy with diffuse shareholders. In the next subsection, we show that dividend taxation has a first-order efficiency cost in such an environment.

5.2 Diffuse Shareholders [$\beta_B < 1$]

Now consider the case where the fraction of shares owned by the board (β_B) is fixed at a level less than 1, so that there are some small shareholders who do not monitor the firm in equilibrium. The manager's objective remains the same as above for given values of $\beta = \alpha(1 - t_d)$ and γ. The board chooses β and γ to maximize its own welfare:

$$W_B = \beta_B(1 - t_d - \beta)P(\gamma, \beta) - c(\gamma) \tag{25}$$

This objective differs from the objective used to choose β and γ in the single shareholder case in only one respect: the board places a weight $\beta_B < 1$ on net profits since it owns only part of the outstanding shares. The small "minority" shareholders are passive, and their welfare depends on the board and manager's choices:

$$W_M = (1 - \beta_B)(1 - t_d - \beta)P(\gamma, \beta)$$

Since the manager's surplus is zero as above, total surplus W in the economy is:

$$W = t_d P + W_B + W_M$$

It is easy to see that $dW_B/dt_d = -\beta_B P$ because of the envelope conditions for β and γ from board maximization. Furthermore, note β maximizes $(1-t_d-\beta)P(\gamma, \beta)$ since c does not vary

[20] Of course, other government regulations and laws can affect the contracting technology in a way that the private sector itself cannot achieve (see e.g, Shleifer and Vishny 1997). For example, if shareholders rights are protected in courts, shareholders may have more control over managers, reducing $c(\gamma)$ and leading to a first-order efficiency gain. The important point is that, keeping constant the regulatory structure embodied by the function $c(\gamma)$, dividend taxes do not affect contracting technology directly.

with β. Hence, the effect of the tax increase on the minority shareholders' utility is:

$$\frac{dW_M}{dt_d} = -(1-\beta_B)P + (1-\beta_B)(1-t_d-\beta)\frac{\partial P}{\partial \gamma}\frac{d\gamma}{dt_d}.$$

Since P is fully determined by β and γ, the last term in this expression is equivalent to the derivative of $P(\gamma, \beta)$ with respect to t_d keeping β constant, which we denote by $dP/dt_d|_\beta$. It follows that

$$\frac{dW}{dt_d} = t_d\frac{dP}{dt_d} + (1-\beta_B)(1-\alpha)(1-t_d)\frac{dP}{dt_d}|_\beta. \tag{26}$$

Equation (26) shows that when shareholders are diffuse, the deadweight cost of raising the dividend tax is the sum of the second-order term obtained with a single owner and a new first-order term proportional to the fraction of small shareholders $(1-\beta_B)$.

The deadweight burden is first-order when $\beta_B < 1$ because the board chooses γ to maximize W_B rather than total shareholder surplus $(W_B + W_M)$, ignoring the spillover benefits of monitoring to the minority shareholders. Since the minority shareholders prefer to free-ride, the equilibrium level of monitoring is suboptimal even relative to the second-best efficient level where total shareholder surplus is maximized net of monitoring costs. This violates condition (ii) in Lemma 2, and effectively creates a public good provision problem where a Pigouvian subsidy to increase supply of the public good (monitoring) would generate a first-order improvement in welfare (see e.g. Kaplow 2006).[21] A dividend tax moves precisely in the opposite direction by reducing the incentive to monitor, leading to a first-order efficiency cost.

In the first-order term in (26), the derivative of P with respect to t_d is taken keeping β constant. This is because the level of β chosen by the board is optimal from the minority shareholders' perspective as well: the level of β that maximizes $\beta_B(1-t_d-\beta)P(\beta,\gamma)$ also maximizes $(1-\beta_B)(1-t_d-\beta)P(\beta,\gamma)$. Intuitively, the board is able to share the cost of increasing the manager's shareownership with the minority shareholders because increasing α dilutes the shareholding of both small and large shareholders proportionally. In contrast, the board cannot share the cost $c(\gamma)$ with the minority shareholders, and therefore sets γ at a suboptimally low level. Thus, only the effect of t_d on P through γ leads to an externality effect.

[21] Condition (i) still holds for the parametrization $x = (\beta, \gamma)$ as $W_S = (1-t_d-\beta)P(\gamma,\beta) - c(\gamma)$.

Equation (26) confirms the critical role of second-best inefficiency of contracts in generating a first-order efficiency cost of taxation. It also explains why formula for deadweight burden differs from that obtained in Gordon and Dietz's (2006) agency model. Gordon and Dietz assume that the board of directors optimizes on behalf of *all* shareholders, i.e. chooses γ as if $\beta_B = 1$. Therefore, there is no free-riding problem in monitoring. The efficiency cost of dividend taxation takes the second-order Harberger form since both conditions in Lemma 2 are satisfied in their model.

An Alternative Representation. In order to implement (26) empirically, one would need to estimate $dP/dt_d|_\beta$, which could be difficult given available data. However, it is straightforward to obtain an alternative representation of the formula that is more convenient from an empirical perspective. Observe that

$$\frac{dP}{dt_d}|_\beta = \frac{dP}{dt_d} - \frac{\partial P}{\partial \beta}\frac{d\beta}{dt_d}.$$

The first order condition with respect to β in the board's objective implies that $(1 - t_d - \beta)(\partial P/\partial \beta) - P = 0$. Recognizing that $\beta = (1 - t_d)\alpha$, it follows that

$$\frac{dP}{dt_d}|_\beta = \frac{dP}{dt_d} + (1 + \varepsilon_{\alpha,1-t_d})\frac{\alpha P}{(1 - t_d)(1 - \alpha)} \tag{27}$$

where $\varepsilon_{\alpha,1-t_d} = [(1 - t_d)/\alpha]d\alpha/d(1 - t_d)$ is the elasticity of manager shareownership w.r.t. the net-of-tax rate. Plugging (27) into (26), we obtain

$$\frac{dW}{dt_d} = -P \cdot \varepsilon_{P,1-t_d}\left[\frac{t_d}{1 - t_d} + (1 - \beta_B)(1 - \alpha)\right] + \alpha \cdot (1 - \beta_B)(1 + \varepsilon_{\alpha,1-t_d}). \tag{28}$$

Equation (28) is easier to implement empirically than (26) because one could in principle estimate the total elasticity $\varepsilon_{P,1-t_d}$ and $\varepsilon_{\alpha,1-t_d}$ by estimating the effect of dividend tax changes on dividend payouts and managerial share ownership. Intuitively, $dP/dt_d|_\beta$ can be inferred by estimating $\varepsilon_{\alpha,1-t_d}$ and subtracting out the effect of the change in α from $\varepsilon_{P,1-t_d}$. We summarize the efficiency consequences of the agency model in the right-hand-side of Table 1.

5.3 Endogenous β_B and Equity Issues

To complete our analysis, we consider a model in which the ownership structure of the firm is fully endogenous, i.e. both α_B and α are determined endogenously. When α_B is endogenous, it is relatively straightforward to allow the firm to issue new equity (E). To make equity issues potentially desirable, we drop the assumption that $f'(X) > 1+r$, leaving X unrestricted. The model we analyze in this section therefore nests both the old and new view models, and the formula for deadweight burden applies to both types of firms. We begin by modelling the determination of E and β_B, and then turn to the efficiency analysis. The main result is that allowing for endogenous α_B and E does not affect the formula for deadweight burden in (26).

Model of E and β_B. We break the manager and shareholder's choices into two stages. In the second stage, the manager chooses D and I conditional on his contract, as in the two cases analyzed above. In the first stage, the external shareholders choose E, and an acquirer buys a fraction β_B of the outstanding shares to take control of the board and set the manager's contract (β, γ). We model the first stage using a Nash equilibrium as follows. First, the dispersed shareholders issue equity E to maximize the value of the firm, taking as given the choices of the board and manager since each small shareholder's equity issue decision has little impact on the firm's overall cash holdings $(X + E)$. The acquirer then makes a tender offer to acquire control by buying a fraction β_B of the company and making a contract (β, γ) with the manager. The acquirer picks (β_B, β, γ) in order to maximize his surplus, taking the equity issue choice E of dispersed shareholders as given.

We begin by characterizing the manager's behavior in the second stage and work backwards. Conditional on β, γ, and $S(\beta, \gamma)$, the firm's manager chooses D and I to maximize

$$\pi u((\beta+\gamma)[D+\frac{1}{1+r}f(I)-S]+S)+(1-\pi)u((\beta+\gamma)[D-S]+S)+\frac{1}{1+r}g(X+E-I-D-S)$$

taking all other variables as fixed. This leads to a function $P(\beta, \gamma, E) = D + \pi f(I)/(1+r) - S$ that gives the total expected payout in terms of the manager's contract and equity raised.

To model how β_B is determined in the first stage, suppose that the company is initially owned by a group of dispersed shareholders. A wealthy shareholder enters the market and buys shares of the company to acquire control of the board through tender offers. This acquiring shareholder starts with no holdings of the company ($\beta_B^0 = 0$). The acquirer buys the shares in bulk at a price equal to the value of the shares after the acquisition (consistent with current practice in tender offers), anticipating the final equilibrium value of the firm.[22]

[22] This model of the formation and value of large block shareholders follows the corporate finance literature, starting with the seminal contribution of Shleifer and Vishny (1986).

Absent any additional benefit from controlling a company, no individual would want to acquire the firm because he must incur the monitoring costs $c(\gamma)$ when he has control of the board. To explain the presence of large shareholders, we introduce a non-market value of controlling a fraction β_B of the company $K(\beta_B)$. The function K captures benefits to board members such as perks or utility of control.[23] We assume that $K(\beta_B)$ has an inverted-U shape – increasing with β_B at low levels, reaching a peak, and then decreasing with β_B. This shape captures the idea that a very low value of β_B does not yield any control, but liquidity constraints or lack of diversification make a very high β_B costly to the shareholder. The inverted-U shape of $K(\beta_B)$ guarantees an interior optimum in β_B.

Taking E and t_d as given, the board of directors chooses β_B, β and γ to maximize

$$\beta_B[(1 - t_d - \beta)P(\gamma, \beta, E) - P_*] - c(\gamma) + K(\beta_B) \tag{29}$$

where P_* is the equilibrium price of the firm, which the acquirer takes as fixed when making the tender offer to the dispersed shareholders. The first order conditions with respect to β, γ, and β_B are:

$$(1 - t_d - \beta)\frac{\partial P}{\partial \beta} = P \tag{30}$$

$$\beta_B(1 - t_d - \beta)\frac{\partial P}{\partial \gamma} = c'(\gamma) \tag{31}$$

$$(1 - t_d - \beta)P(\gamma, \beta, E) - P_* = -K'(\beta_B) \tag{32}$$

In equilibrium, $P_* = (1 - t_d - \beta)P(\gamma, \beta, E)$. Hence, the first order condition for β_B simplifies to $K'(\beta_B) = 0$, which implies that β_B is independent of t_d and E in equilibrium. The solution of the acquirer's problem thus yields functions $\beta(t_d, E)$, $\gamma(t_d, E)$ and a constant β_B.

Finally, the dispersed shareholders choose E to maximize their total dividend payout net of equity investment. We assume that dispersed shareholders anticipate the equilibrium levels of β and γ after the takeover when choosing E. However, the dispersed shareholders do not

[23] In the appendix, we show that modelling such benefits as kickbacks from the manager to the large shareholder (a "corrupt board") yields similar results.

internalize the effect of their choice of E on β and γ because they are small. Hence, E is chosen to maximize

$$W_M = (1 - t_d - \beta)P(\beta, \gamma, E) - E \tag{33}$$

The first order condition with respect to E is:

$$(1 - t_d - \beta)\frac{\partial P}{\partial E} = 1, \tag{34}$$

The solution to (34) yields a function $E(t_d, \beta, \gamma)$. The Nash equilibrium levels of β, γ, and E are given by a triplet (β_*, γ_*, E_*) such that each player's behavior is optimal given the other's choice:

$$\beta_* = \beta(t_d, E_*),\ \gamma_* = \gamma(t_d, E_*),\ E_* = E(t_d, \beta_*, \gamma_*).$$

The equilibrium triplet is a function of t_d, the only remaining exogenous variable in the model.

Efficiency Analysis. Let the equilibrium value of the firm be denoted by

$$V(t_d) = (1 - t_d - \beta_*(t_d))P(\gamma_*(t_d), \beta_*(t_d), E_*(t_d))$$

Since the manager is held to the participation constraint as above, social surplus is

$$W = t_d P + V(t_d) - E - K(\beta_B) - c(\gamma) \tag{35}$$

We can now state our general formula for the efficiency cost of dividend taxation.

Proposition 3 *With endogenous ownership structure and equity issues, excess burden is*

$$\frac{dW}{dt_d} = t_d\frac{dP}{dt_d} + (1 - \beta_B)(1 - \alpha)(1 - t_d)\frac{dP}{dt_d}|_{\beta,E}. \tag{36}$$

Proof. Totally differentiating V with respect to t_d gives

$$V'(t_d) = -P + \left[-P + (1 - t_d - \beta)\frac{\partial P}{\partial \beta}\right]\beta_*'(t_d) + (1 - t_d - \beta)\frac{\partial P}{\partial \gamma}\gamma_*'(t_d) + (1 - t_d - \beta)\frac{\partial P}{\partial E}E_*'(t_d).$$

Using the first order conditions (30) and (34), this expression simplifies to:

$$V'(t_d) = -P + (1 - t_d - \beta)\frac{\partial P}{\partial \gamma}\gamma_*'(t_d) + E_*'(t_d)$$

Next, using (31), differentiating (35) implies that

$$\begin{aligned}\frac{dW}{dt_d} &= t_d\frac{dP}{dt_d} + P + V'(t_d) - E'_*(t_d) - c'(\gamma)\gamma'_*(t_d) \\ &= t_d\frac{dP}{dt_d} + (1-\beta_B)(1-t_d-\beta)\frac{\partial P}{\partial\gamma}\gamma'_*(t_d),\end{aligned}$$

which can be rewritten as (36). QED.

Equation (36) is the same formula that we obtained in equation (26) with fixed β_B and no equity issues, except that both β and E are held constant when computing $\frac{dP}{dt_d}$ in the first-order term in (36).[24] Intuitively, E is set at the second-best efficient level that maximizes total private surplus because shareholders, and not entrenched managers, choose E. When choosing the amount of equity, the dispersed shareholders internalize the benefits that accrue to other shareholders (e.g. the block holder) because of the ability to trade shares after the equity issue. The optimality of the equity decision explains why the first-order term in (36) depends on dP/dt_d keeping both β and E fixed, as there is no externality problem in the choice of these variables.

Allowing for endogenous determination of β_B does not affect the formula in (26) at all because the acquirer chooses β_B to maximize his *private* rather than social surplus (neglecting the minority shareholders). The free-rider externality problem therefore remains unresolved. As a result, dividend taxation leads to a first-order efficiency cost by distorting the choices of β_B and γ. Both of these channels are taken into account in the empirical estimate of $\frac{dP}{dt_d}|_{\beta,E}$, leaving the formula for $\frac{dW}{dt_d}$ unchanged.

It is instructive to compare the effects of making β_B (board shareownership) and α (manager shareownership) endogenous. As discussed in section 5.1, when $\beta_B = 1$, deadweight burden is first-order when α is fixed but second-order when α is endogenous. Allowing for endogenous determination of α makes deadweight burden second-order because the costs of raising α are shared through dilution, and thus α is effectively chosen to maximize *total* private surplus even with multiple shareholders. In contrast, deadweight burden remains first-order when β_B is endogenous because β_B is not chosen to maximize total private surplus, since the costs of controlling and monitoring the firm are borne only by the large shareholder. This

[24] The value of $\frac{dP}{dt_d}|_{\beta,E}$ can be estimated empirically using the same method as in the second case above. In particular, an estimate of $\frac{\partial E_*(t_d)}{\partial t_d}$ can be used to remove the E channel from dP/dt_d using the first-order condition in (34), just as we removed the β channel in (28).

contrast between the effects of making α and β_B endogenous underscores the central role of the second-best efficiency of private contracts in determining the efficiency cost of taxation.

5.4 Extensions

In the appendix, we analyze two additional extensions of the model to evaluate the robustness of the formulas derived above. First, we extend the model of Section 5.3 to allow the firm to raise funds through debt (L) in addition to equity E. Debt pays a fixed interest rate r and also differs from equity in its tax treatment, in that interest payments are not subject to the dividend tax. The corporate finance literature has emphasized that debt finance is more costly than equity because of the risk of an expensive bankruptcy, which explains why companies use equity despite the tax advantage of debt (see e.g., Jensen and Meckling 1976, Brealey and Myers 2003). We model the cost of carrying debt in reduced-form through a convex cost function $c_L(L)$. The dispersed shareholders choose both E and L in the first stage; all other stages are as above. We obtain the same formula for $\frac{dW}{dt_d}$ as in Proposition 3, except that the first-order term depends on $\frac{dP}{dt_d}|_{\beta,E,L}$ for reasons analogous to those discussed above with endogenous equity issues.

Second, we extend the model of Section 5.2 to allow for an internal "culture of corruption" within the board of directors. Recent studies have argued that the board itself may receive implicit or explicit private benefits from the manager's pet projects (see e.g., Shleifer and Vishny 1997). Such transfers add to the benefits of control for the acquirer (K), but create a further disincentive to monitor the manager. We model such corruption by adding a term $\sigma \cdot g(J)$ to the board of director's welfare in (25). In this case, the formula in (28) carries over without any changes.[25] The efficiency cost formulas are generally robust because introducing additional choice variables or constraints into the manager's and shareholders' problems does not affect the key envelope conditions, as long as these choices do not have externalities on other agents.

Following standard practice in the public finance literature, we have ignored share repurchases as a means of returning money to shareholders. Unlike the other simplifying assumptions, the exclusion of repurchases is a substantive limitation of our analysis. Share

[25] Formula (26) with $dP/dt_d|_\beta$ cannot be used when $\sigma > 0$ because the corruption term $\sigma g(J(\beta,\gamma))$ introduces an externality in the choice of β as well, which has to be taken into account in the marginal deadweight burden computation.

repurchases can affect the efficiency cost of dividend taxation in different ways, depending on how the repurchase decision is modelled. If repurchases are effectively negative equity issues, our formula is unchanged since the analysis above permits negative E. This assumes, however, that the repurchase decision is made by shareholders rather than the manager. If the repurchase decision is made by the manager, and capital gains and dividends are taxed at the same rate, then (36) can still be used, replacing dividend payouts with the sum of dividends and repurchases. If the tax rates differ, the efficiency cost is more complicated, and will in general depend on both the repurchase and dividend elasticities. Both payout choices create externalities on the dispersed shareholders and therefore create first-order distortions, potentially of opposite sign. In view of the sensitivity of the results to the way in which repurchases are modelled, we do not attempt to characterize the efficiency cost of dividend taxation in an environment with repurchases here.

6 Illustrative Calibration

A useful feature of the formula in Proposition 3 is that it depends on a small set of parameters that can in principle be estimated empirically. The primitives of the model – e.g. the monitoring and share acquisition cost functions $c(\gamma)$ and $K(\beta_B)$ or the pet project payoff $g(J)$ – affect efficiency costs only through the high-level elasticities and equilibrium ownership structure (e.g. $\varepsilon_{P,1-t}$, α, α_B). Since these high-level inputs are estimated empirically, our method of computing the efficiency cost of taxation is robust to variations in the structural parametrization of the model. This is important because estimating the deep structural parameters would be difficult, especially since they represent reduced forms of complex contracts and payoffs for shareholders and management.

As an illustration, we calibrate the marginal deadweight cost of raising the dividend tax from the current rate of 15% relative to the marginal revenue from raising the tax. We assume that equity and debt issues do not vary with t in this calculation, so that $\frac{dP}{dt_d}|_{\beta,E,D} = \frac{dP}{dt_d}|_{\beta}$. In this case, the formula in Proposition 3 coincides with the empirically convenient representation in (28). Note that managerial share ownership (α) is quite small relative to $\varepsilon_{P,1-t_d}$ in practice: in the Execucomp sample analyzed in Figure 1, total executive shareownership averages less than $\alpha = 0.03$ in all years.[26] Therefore, the second term in (28), which is proportional to α,

[26] Although our calculation focuses solely on stock ownership, accounting for other forms of incentive-based

is likely to be negligible in magnitude. As an approximation we ignore this term and assume $\alpha = 0$. This simplifies (28) to:

$$\frac{dW}{dt_d} = -P\varepsilon_{P,1-t_d}\left[\frac{t_d}{1-t_d} + (1-\alpha_B)\right] \tag{37}$$

The marginal revenue from raising the dividend tax rate is

$$\frac{dR}{dt_d} = P\left[1 - \frac{t_d}{1-t_d}\varepsilon_{P,1-t_d}\right] \tag{38}$$

Although we interpreted α_B as "board shareownership" in the model, large shareholders could potentially influence manager's decisions even if they are not officially on the board. We therefore calibrate α_B using a broader measure by including both board members and all other large (5%+) blockholders. Using data from the IRRC and Dlugosz et al. (2007), we estimate that (non employee) board members and large blockholders together owned an average of $\alpha_B = 0.1$ of shares between 1998 and 2001 (see data appendix for details).[27] Plugging $\alpha_B = 0.1$, $\varepsilon_{P,1-t_d} = 0.75$ (Chetty and Saez 2005), and $t_d = 0.15$ into (37) and (38), we obtain

$$\frac{dW}{dt_d}/\frac{dR}{dt_d} = -0.93.$$

Generating \$1 of additional revenue from the dividend tax starting from a rate of 15% would generate deadweight loss of 93 cents in the short run. 15 cents of this deadweight burden comes from the conventional Harberger effect and 78 cents comes from the first-order amplification of the free-rider distortion in the agency model.

As a different method of gauging the magnitude of the efficiency cost, we calculate the cost of increasing the dividend tax rate back to the pre-2003 level. The 2003 dividend tax reform cut the tax rate on qualified dividends paid to individuals by about 20 percentage points. Qualified dividend payments to individuals totalled $P =$ \$119 billion in 2005 (US Treasury Department: Internal Revenue Service, 2007). Using the linear approximation of $W(t_d)$ in (37), a 20 percentage point increase in the dividend tax rate would reduce total social welfare by the equivalent of \$19 billion, or 0.2% of GDP. Equation (38) implies that the tax increase would produce approximately \$21 billion of additional government revenue in the short run.

pay is unlikely to raise α significantly. Existing studies have measured α more broadly by computing the change in the wealth of a CEO when his firm's value increases by \$1. These studies estimate that α is less than 1% on average for CEOs of publicly traded corporations in the U.S. (see Murphy 1999 for a survey).

[27] Unfortunately, we could not study the effect of the 2003 dividend tax reform on the broader measure of α_B in Figure 1 because the Dlugosz et al. data on large blockholders is available only through 2001.

These calculations should be viewed as illustrative because they ignore several potentially important issues. First, and most importantly, the estimate of $\varepsilon_{P,1-t_d}$ in Chetty and Saez (2005) is a short-run elasticity: it compares changes in dividend payments in a three-year window around the 2003 reform. The long-run elasticity could be significantly smaller if companies paid out excess cash holdings that they had built up immediately after the tax cut. In addition, as shown in Figure 1, the ownership structure of the firm appears to have remained roughly constant over the horizon we studied. In the longer run, α and E would respond to the tax regime, further affecting $\varepsilon_{P,1-t_d}$ and introducing a wedge between dP/dt_d and $dP/dt_d|_{\beta,E}$. A second limitation of our calibration is that it ignores share repurchases. In Chetty and Saez (2006), we present suggestive evidence that companies did not substitute dividends for repurchases, but further analysis is needed on this issue. Third, our analysis ignores the potential role of dividends as signals, in which case dividend taxation could have little impact on efficiency despite reducing dividend payments (Bernheim 1991). Finally, as in any normative analysis, one must consider the distributional implications of dividend taxation in addition to its efficiency consequences. If the incidence of dividend taxes are borne by wealthy shareholders, it is possible that taxing dividends does not have a large impact on social welfare from a utilitarian perspective.

7 Conclusion

The public finance literature on corporate taxation is based primarily on models of profit-maximizing firms. In contrast, since Jensen and Meckling (1976), the corporate finance literature has emphasized deviations from profit maximization by managers as a central determinant of firm behavior. This paper has taken a step toward bridging this gap. We analyzed the effects of dividend taxation in a standard agency model, and showed that it can explain many aspects of the empirical evidence that pose problems for existing models.

The agency model yields several new insights into the efficiency costs of dividend taxation. Efficiency costs depend critically on the ownership structure of the firm. If the firm is owned by a single principal who monitors the manager, dividend taxation has a second-order efficiency cost. Despite overinvestment due to the agency problem, the Harberger formula applies because the private market still achieves a second-best efficient outcome given the informational constraints inherent in the manager-shareholder relationship. However, if the firm is owned by many shareholders, as are most modern corporations, the private market is unlikely to achieve a second-best efficient outcome because of free-riding in monitoring. As a result, Pareto improvements which respect the informational constraints are left unexploited in the equilibrium contract between shareholders and managers. In this environment, dividend taxation exacerbates the under-monitoring problem and creates a first-order efficiency loss.

An illustrative calibration of our formula for deadweight cost using empirical estimates from the 2003 tax reform points to two general lessons. First, the main potential source of inefficiency from increasing the dividend tax rate is the misallocation of capital by managers because of reduced monitoring, and not the classic distortion to the overall level of investment emphasized in the "old view" model. Second, from a policy perspective, our analysis suggests that dividend taxation should be used relatively little given its inefficiency if the government has other tools (e.g. progressive income taxation integrated with corporate taxation) that have similar distributional effects but do not create first-order distortions. In fact, a dividend subsidy may be desirable to subsidize monitoring and correct the free-rider problem.

Our analysis calls for further empirical work related to agency issues in corporate taxation. In our model, a dividend tax cut raises efficiency by improving the allocation of capital: firms with excess cash holdings invest *less* following a tax cut, while cash-constrained firms invest more. Testing whether tax reforms generate such heterogeneous investment responses across firms would shed light on the empirical importance of this allocation efficiency mechanism.

Our analysis can also be refined in several respects in future theoretical work. Although our formula for deadweight burden generalizes the existing "old view" and "new view" results, it does not nest the signalling model. The signalling model can explain why companies pay dividends despite the tax advantage of repurchases (Bernheim 1991), unlike our simple agency model. A model that combines signalling and agency effects could therefore yield a more precise understanding of the efficiency costs of dividend taxation in the presence of share repurchases.[28] More generally, further research on the interactions between agency theory and taxation could yield insights into the welfare consequences of other important policy issues such as the corporate tax, the capital gains tax, and tax treatment of interest payments.

[28] Such an analysis would be particularly interesting in view of Bernheim's (1991) striking result that the efficiency cost of taxation is zero even though dividends fall with the tax rate in the signalling model.

Appendix A: Extensions of Efficiency Analysis

Corrupt Board.

Consider the model in section 5.2 with fixed β_B for simplicity; as above, allowing for endogenous determination of β_B does not affect the results. To allow for imperfect monitoring of the managers by the board because of internal corruption, we introduce an additional term $\sigma \cdot g(J)$ into the board's welfare. This changes the board's objective when setting α and γ to

$$W_B = \beta_B(1 - t_d - \beta)P(\gamma, \beta) - c(\gamma) + \sigma g(J(\gamma, \beta))$$

The manager's problem and dispersed shareholders' welfare are the same as in section 5.2 conditional on β and γ. Total surplus is $t_d P + W_B + W_M$. As is now familiar, we exploit the envelope conditions for γ, and β in W_B to obtain:

$$\frac{dW}{dt_d} = P + t_d \frac{dP}{dt_d} - \beta_B P - (1 - \beta_B)P + (1 - \beta_B)(1 - t_d - \beta)\frac{dP}{dt_d} - (1 - \beta_B)P\frac{d\beta}{dt_d}$$

which again coincides with the formula for deadweight burden in (28). One way to understand this result is by viewing the $\sigma g(J(\gamma, \beta))$ term as a modification of the cost function $c(\gamma)$. Since c does not enter our formula for deadweight burden, allowing for internal corruption via $\sigma > 0$ does not affect the results.

Debt Finance.

Suppose the firm can raise funds from both the stock market via new equity issues (E) and the bond market by issuing debt ($L \geq 0$), which pays a fixed return r that is deducted from profits. Issuing debt has a cost $c_L(L)$, e.g. reflecting the risk of bankruptcy, which is deducted from the firm's expected profits. The firm's total cash available for dividends and investment is $X + E + L$ but the firm has to repay $(1 + r)L$ at the end of period 1 (i.e., has to repay L from a period 0 perspective). The firm's and shareholder's choices can be broken into two stages as in section 5.3. First, there is a game where dispersed shareholders issue debt and equity (L, E) and the board of directors forms through costly acquisitions of a large block of shares (β_B) and chooses a contract α, γ with the manager. Second, the manager chooses D and I given the contract he faces.

Conditional on β, γ, E, L the firm's manager chooses D and I to maximize

$$\pi u((\beta+\gamma)[D+\frac{1}{1+r}f(I)-S-L]+S)+(1-\pi)u((\beta+\gamma)[D-S-L]+S)+\frac{1}{1+r}g(X+E+L-I-D-S)$$

In this setting, the firm's total expected dividend payout is

$$P(\beta,\gamma,E,L) = D(\beta,\gamma,E,L) + \pi f(I(\beta,\gamma,E,L))/(1+r) - S(\beta,\gamma,E,L) - L - c_L(L)$$

As in the text, conditional on E, L, and t_d the board of directors chooses β, γ, and β_B to maximize

$$W_B = \beta_B[(1-t_d-\beta)P(\beta,\gamma,E,L) - V(t_d)] - c(\gamma) - K(\beta_B)$$

The first order condition for β_B is again $K'(\beta_B) = 0$ which shows that β_B is independent from other variables.

Finally, the dispersed shareholders choose E and L to maximize their total payout net of equity investment conditional on β, γ, and t_d:

$$W_M = (1-t_d-\beta)P(\beta,\gamma,E,L) - E$$

The maximization programs lead to a Nash equilibrium quartuple $(\beta_*, \gamma_*, E_*, L_*)$ which depends only on t_d. As above, we denote the net value of the firm in equilibrium by $V(t_d) = (1-t_d-\beta)P(\beta,\gamma,E,L)$.

Social surplus is defined as

$$W = t_d P + W_B + W_M = t_d P(t_d) + V(t_d) - E_*(t_d) - c(\gamma_*(t_d)) - K(\beta_B)$$

as in the text. Because L maximizes $P(\beta,\gamma,E,L)$, $\partial P/\partial L = 0$. Following the same steps as in the proof of Proposition 3 yields the same expression as (36), with $\frac{dP}{dt_d}|_{\beta,E}$ replaced by $\frac{dP}{dt_d}|_{\beta,E,L}$.

Appendix B: Data for Figure 1 and Table 2

Data Sources.

Executive Share Ownership. Data on top executive ownership comes from the Compustat Executive Compensation (ExecuComp) database. For each firm/year, executive share ownership is calculated as the total number of shares excluding options held by executives covered by ExecuComp divided by the total number of shares outstanding. The firm-level data is aggregated by taking a weighted mean across all firms, with (fixed) weights equal to the average market capitalization (reported in ExecuComp) between 1996 and 2005.

Board of Director Share Ownership. Data on director ownership comes from the Investor Responsibility Research Center (IRRC) Director database, available through Wharton Research Data Services. For each firm/year, director share ownership is calculated as the total number of shares held by non-employee directors (as identified by the variable DIRTYPE) divided by the total number of shares outstanding from ExecuComp. To reduce measurement error, a small number of observations are omitted if the implied percentage of director ownership exceeds 100%. As with executive ownership, the annual series is computed as a weighted mean with weights equal to the average market capitalization between 1996 and 2005.

Dividend Payments. Data on dividend payments come from the Center for Research in Security Prices (CRSP). The annual series is calculated as the sum of all taxable regular (non special) dividends paid by U.S. corporations (excluding utilities and financials) listed on the NYSE, AMEX, and NASDAQ stock exchanges. See Chetty and Saez (2005) for more details.

Calibration of α_B in Section 6.

The value for α_B used in the calibration is based on the sum of the average fraction of shares held by non-employee directors and the average fraction held by other large independent blockholders (defined as holdings that exceed 5% of the total shares outstanding). The data on large blockholders comes from the dataset constructed by Dlugosz et al. (2007), available through WRDS. The data on board ownership is from the raw IRRC dataset described above.

The average level of non-employee director ownership in the IRRC between 1998 and 2001 is 2.5%. The average level of outside blockholder ownership (as defined by the variable SUMOUT) between 1998 and 2001 (the last year of data) for the same set of firms in the Dlugosz et al. dataset is 7.5%. In both calculations, the average is calculated as the mean weighted by each firm's average market capitalization over the sample period. Summing these two estimates, we obtain $\alpha_B = 10\%$.

References

Allen, Franklin, Antonio Bernardo and Ivo Welch, 2000, "A Theory of Dividends Based on Tax Clientele" *Journal of Finance* 55(6), 2499-2536.

Allen, Franklin and Roni Michaely, 2003, "Payout Policy," in Constantinides, George, Milton Harris and Rene Stulz eds. *Handbook of the Economics of Finance* (Amsterdam: North-Holland)

Auerbach, Alan, 1979, "Wealth maximization and the cost of capital," *Quarterly Journal of Economics* 93, 433-46.

Auerbach, Alan (1985), "The Theory of Excess Burden and Optimal Taxation," in Auerbach, Alan and Martin Feldstein eds. *Handbook of Public Economics*, Volume 1, 61-127, (Amsterdam: North-Holland)

Auerbach, Alan, 2003, "Taxation and Corporate Financial Policy," in Auerbach, Alan and Martin Feldstein eds. *Handbook of Public Economics*, Volume 3, (Amsterdam: North-Holland)

Auerbach, Alan and Kevin Hassett, 2002, "On the Marginal Source of Investment Funds," *Journal of Public Economics* 87(1), 205-232.

Auerbach, Alan and Kevin Hassett, (2007) "The 2003 Dividend Tax Cuts and the Value of the Firm: An Event Study," in A. Auerbach, J. Hines, and J. Slemrod, eds., *Taxing Corporate Income in the 21st Century*, (Cambridge: Cambridge University Press), 93–126.

Auerbach, Alan and James Hines (2003), "Taxation and Economic Efficiency," in Auerbach, Alan and Martin Feldstein eds. *Handbook of Public Economics*, Volume 3, (Amsterdam: North-Holland)

Bernheim, B. Douglas 1991, "Tax Policy and the Dividend Puzzle," *RAND Journal of Economics* 22, 455-76.

Bernheim, B. Douglas and Lee S. Redding, 2001. "Optimal Money Burning: Theory and Application to Corporate Dividends," *Journal of Economics and Management Strategy* 10, 463-507.

Bernheim, B. Douglas and Adam Wantz (1995), "A Tax-Based Test of the Dividend Signalling Hypothesis", *American Economic Review* 85(3), 532-551.

Bertrand, Marianne and Sendhil Mullainathan (2003), "Enjoying the Quiet Life? Corporate Governance and Managerial Preferences," *Journal of Political Economy* October.

Bradford, David F., 1981, "The incidence and allocation effects of a tax on corporate distributions," *Journal of Public Economics* 15, 1-22.

Brealey, Richard and Stuart Meyers (2003), *Principles of Corporate Finance,* Mc-Graw Hill.

Brown, Jeffrey, Nellie Liang, and Scott Weisbenner (2006) "Executive Financial Incentives and Payout Policy: Firm Responses to the 2003 Dividend Tax Cut", forthcoming in *Journal of Finance.*

Chetty, Raj and Emmanuel Saez (2005), "Dividend Taxes and Corporate Behavior: Evidence from the 2003 Dividend Tax Cut", *Quarterly Journal of Economics* 120(3), 791-833.

Chetty, Raj and Emmanuel Saez (2006), "The Effects of the 2003 Dividend Tax Cut on Corporate Behavior: Interpreting the Evidence", *American Economic Review Papers and Proceedings* 96(2), 124-129.

Chirinko, Robert S. 1993, "Business Fixed Investment Spending: Modeling Strategies, Empirical Results, and Policy Implications" *Journal of Economic Literature*, 31(4), 1875-1911.

Christie, William and Vikram Nanda, 1994, "Free Cash Flow, Shareholder Value, and the Undistributed Profits Tax of 1936 and 1937," *Journal of Finance*, 49 (5) 1727-1754.

Coase, Ronald, 1960, "The Problem of Social Cost," *Journal of Law and Economics* 3(1), 1-44.

Desai, Mihir A., C. Fritz Foley, and James Hines (2007), "Dividend Policy inside the Multinational Firm," NBER Working Paper No. 8698, forthcoming in *Financial Management*.

Desai, Mihir A. and Austan D. Goolsbee, 2004 "Investment, Overhang and Tax Policy" *Brookings Papers on Economic Activity* 2, 285-355.

Dlugosz, Jennifer, Rudiger Fahlenbrach, Paul Gompers, and Andrew Metrick, (2007), "Large Blocks of Stock: Prevalence, Size, and Measurement", *Journal of Corporate Finance*, forthcoming.

Easterbrook, Frank H., 1984, "Two Agency-Cost Explanations of Dividends," *American Economic Review*, 74 (4), 650-659.

Feldstein, Martin, (1970) "Corporate taxation and Dividend Behavior," *Review of Economic Studies* 37, 57-72.

Fenn, George W. and Nellie Liang, 2001, "Corporate payout policy and managerial stock incentives", *Journal of Financial Economics*, 60(1), 45-72.

Gordon, Roger and Martin Dietz, 2006, "Dividends and Taxes", NBER Working Paper No. 12292, forthcoming in eds. Alan Auerbach and Daniel Shaviro *Institutional Foundations of Public Finance: Economic and Legal Perspectives*, (Harvard University Press, Cambridge MA).

Goulder, Lawrence and Roberton Williams (2003) "The Substantial Bias from Ignoring General Equilibrium Effects in Estimating Excess Burden, and a Practical Solution," *Journal of Political Economy*, 111, 898-927.

Grossman, Sanford J. and Oliver D. Hart, 1980, "Takeover bids, the free-rider problem, and the theory of the corporation", *Bell Journal of Economics* 11, 42-54.

Harberger, A.C., 1962, "The incidence of the corporation income tax", *Journal of Political Economy* 70, 215-240.

Harberger, A.C., 1966, "Efficiency effects of taxes on income from capital" in: M. Krzyzaniak, ed., *Effects of corporation income tax* (Wayne State University Press, Detroit).

Hines, James R., Jr. 1996. "Dividends and Profits: Some Unsubtle Foreign Influences." *Journal of Finance* 51, 661-689.

Hines, James R., Jr. 1999. "Three Sides of Harberger Triangles," *Journal of Economic Perspectives* 13 (2), 167-188.

Jensen, Michael C, 1986, "Agency costs of free cash flow, corporate finance and takeovers," *American Economic Review* 76, 323-329.

Jensen, Michael C. and William H. Meckling , 1976, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, 3(4), 305-360.

John, Kose and Joseph Williams, 1985, "Dividends, Dilution, and Taxes: A Signaling Equilibrium," *Journal of Finance* 40 (4), 1053-1070.

Kaplow, Louis, 2006, "Optimal Control of Externalities in the Presence of Income Taxation", NBER Working Paper No. 12339.

Kaplow, Louis, 2008. *The Theory of Taxation and Public Economics* (Princeton University Press: New Jersey).

King, Mervin A., 1977, *Public Policy and the Corporation,* (Chapman and Hall, London).

Korinek, Anton and Joseph Stiglitz (2006), "Dividend Taxation and Intertemporal Tax Arbitrage", Columbia University Working Paper.

LaPorta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny, 2000, "Agency problems and dividend policies around the world," *Journal of Finance* 55(1), 1-33.

Murphy, Kevin J., 1999, "Executive Compensation", in Ashenfelter, Orley and David Card eds. *Handbook of Labor Economics*, Volume 3, (Amsterdam: North-Holland)

Nam, Jouahn, Jun Wang, and Ge Zhang (2004), "The Impact of Dividend Tax Cut and Managerial Stock Holdings on Firm's Dividend Policy", unpublished mimeograph, Baruch College.

Poterba, James, 2004, "Corporate Payout Policy", *American Economic Review Papers and Proceedings*, 94(2), 171-175.

Poterba, James and Lawrence Summers, 1985 "The Economic Effects of Dividend Taxation", in E. Altman and M. Subrahmanyam, eds., *Recent Advances in Corporation Finance* (Homewood, IL: Dow Jones-Irwin), 227-284.

Prescott, Edward C. and Robert M. Townsend, 1984a "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard" *Econometrica* 52(1): 21-45.

Prescott, Edward C. and Robert M. Townsend, 1984b. "General Competitive Analysis in an Economy with Private Information" *International Economic Review* 25(1), 1-20.

Rajan, Raghuram, Henri Servaes, and Luigi Zingales, 2000, "The cost of diversity; The diversification discount and inefficient investment," *Journal of Finance* 55, 35-80.

Scharfstein, David and Jeremy Stein, 2000, "The Dark Side of Internal Capital Markets: Divisional Rent-Seeking and Inefficient Investment," *Journal of Finance*, 55 (6).

Shleifer Andrei and Robert Vishny, 1986, "Large shareholders and Corporate Control," *Journal of Political Economy* 94 (3), 461-488.

Shleifer, Andrei and Robert Vishny, "A Survey of Corporate Governance," Nobel Symposium on Law and Finance, August, 1995, *Journal of Finance*, June 1997.

Sinn, Hans-Werner, (1991) "The Vanishing Harberger Triangle," *Journal of Public Economics*, 45(3), 271-300.

U.S. Treasury Department, Internal Revenue Service, 2007 *Statistics of Income: Individual Income Tax Returns for Year 2005.* (Government Printing Press: Washington, D.C.).

Table 1

Summary of Key Predictions: Neoclassical vs. Agency Models

	Neoclassical Model			Agency Model	
	Old View		*New View*		
Initial Cash X	Low: $f'(X)>(1+r)/(1-t_d)$	Medium: $(1+r)/(1-t_d)>=f'(X)>=1+r$	High: $1+r>f'(X)$	High: $1+r>f'(X)$ and $g'(X-I^S)>=\omega\ (1+r)$	Very High: $1+r>f'(X)$ and $g'(X-I^S)<\omega\ (1+r)$
Dividends D	D=0	D=0	D>0, $f'(X-D)=1+r$	D=0	D>0, $g'(X-I^S-D)=\omega(1+r)$
Equity Issues E	E>0, $f'(X+E)=(1+r)/(1-t_d)$	E=0	E=0	E=0	E=0
Productive Investment I	I>X, $f'(I)=(1+r)/(1-t_d)$	I=X, $f'(I)=f'(X)$	I<X, $f'(I)=1+r$	$g'(X-I)=\omega f'(I)$, $f'(I)>1+r$	$f'(I)=1+r$, $g'(J)=\omega(1+r)$
Effects of reducing dividend tax t_d	No effect on D I increases, E increases	Intensive margin: No effect on D, E, I Extensive margin: Some firms shift to low cash regime, start issuing E and increase I	No effect on D, E, I	Intensive margin: No effect on D and E, I increases, J decreases Extensive margin: Some firms shift to very high cash regime, start paying dividends	D increases, J decreases No effect on I and E
Heterogeneity of D response to tax cut by ownership structure	None	None	None	Extensive margin: higher likelihood and larger D initiations if exec. or board share high	Larger increase in D if exec. or board share high (if third derivatives of g, c small)
Efficiency cost of t_d	Harberger	Harberger (along extensive margin)	None	Harberger if $\beta_B=1$ Harberger + first order effect if $\beta_B<1$	Harberger if $\beta_B=1$ Harberger + first order effect if $\beta_B<1$

NOTE-- This table summarizes the firm's choice of dividends (D), equity issues (E), and investment (I) in the neoclassical and agency models. Behavior depends on the level of initial cash holding X, which varies across the columns. I^S denotes the efficient investment level such that $f'(I^S)=1+r$. In the agency model, shareholders issue equity if cash holdings are sufficiently low. In this case, the efficiency costs of t_d remain the same as those listed in the last row of the table.

TABLE 2

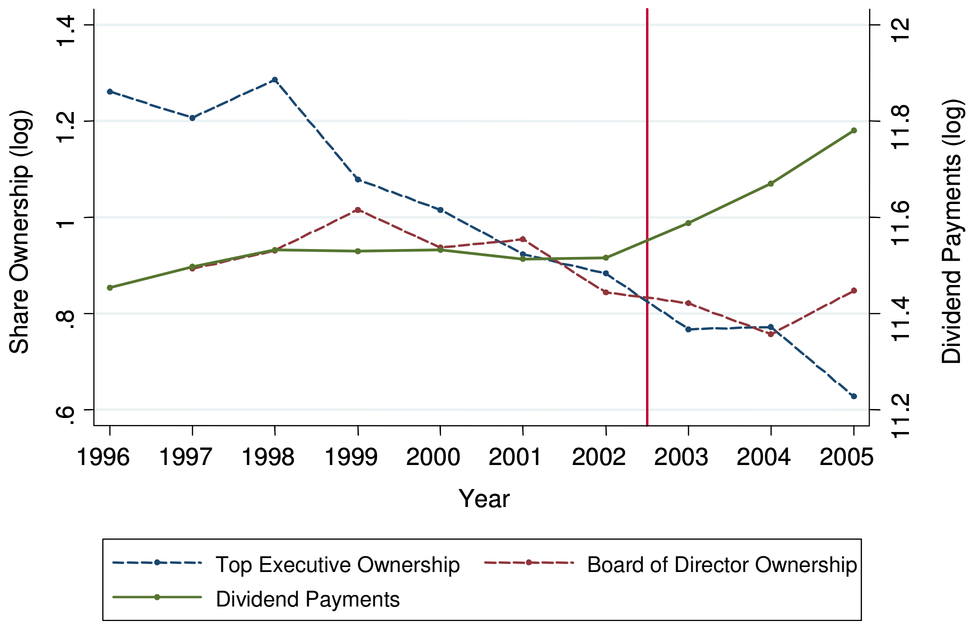
Regression Estimates: Effect of 2003 Tax Cut
on Ownership Structure and Dividend Payments

	(1) Log Executive Ownership Share	(2) Log Director Ownership Share	(3) Log Dividend Payments
Post-2003 Indicator	**0.048**	**-0.048**	**0.148**
	(0.063)	(0.075)	(0.043)
Year	-0.083	-0.014	-0.014
	(0.010)	(0.012)	(0.043)
Constant	166.061	29.367	40.246
	(20.129)	(24.173)	(13.897)
Number of Obs.	9	9	9

NOTE--Each column reports estimates of an OLS regression; standard errors in parentheses. Post-2003 dummy is an indicator for year>=2003. To construct executive shareownership measure, we first compute the total fraction of shares held by executives of each firm listed in Execucomp, and then calculate an annual mean, weighting each firm by its average market capitalization over the sample period. Board ownership share is constructed analogously using data from the Investor Responsibility Research Center. Dividend payments are defined as total dividend payments by all firms listed in CRSP, as in Chetty and Saez (2006). Sample for all regressions includes the years from 1996-2005. See data appendix for additional details on sample and variable definitions.

Figure 1

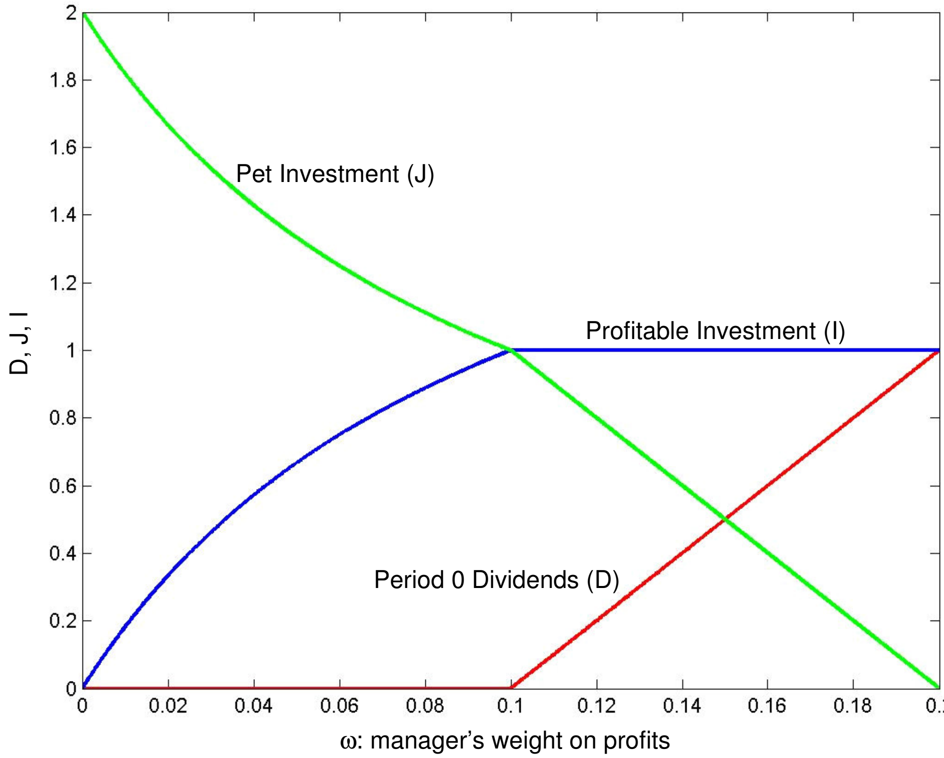
Short-Run Effect of 2003 Tax Reform on Ownership Structure and Dividends



NOTE–This figure plots the annual evolution of three variables around the 2003 dividend tax cut, demarcated by the vertical line. The first series is log average executive shareownership. To construct this measure, we first compute the total fraction of shares held by executives of each firm listed in Execucomp, and then calculate an annual mean, weighting each firm by its average market capitalization over the sample period. Average board of director shareownership is constructed analogously using data for all firms listed in the Investor Responsibility Research Center database. Dividend payments are defined as total regular dividend payments by all firms listed in CRSP, as in Chetty and Saez (2006). The scale for dividend payments (right axis) is chosen so that the range of the right and left axis is the same (0.8 log points) for comparability of trends. See data appendix for additional details on sample and variable definitions.

Figure 2

Manager's Decision Rules as a Function of Weight on Profits



NOTE–This figure plots the manager's optimal choice of dividends, profitable investment, and pet project investment as a function of his weight on profits, ω. The simulation assumes a total cash holding of $X = 2$, profitable investment production function $f(I) = 2I - \frac{I^2}{2}$, pet production function $g(J) = \frac{1}{10}[2J - \frac{J^2}{2}]$, and interest rate $r = 0$.

Figure 3a

Effect of Tax Cut on Dividends by Managerial Shareownership

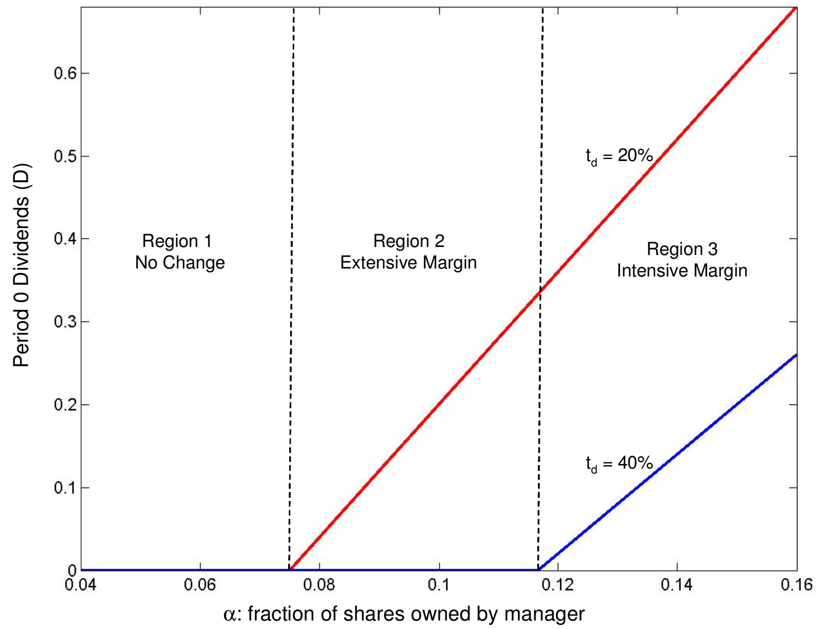
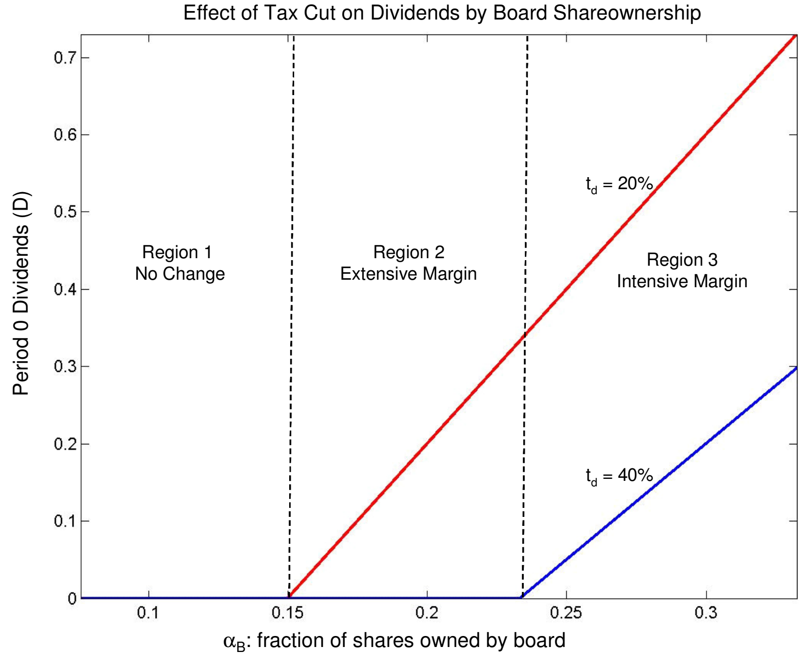


Figure 3b

Effect of Tax Cut on Dividends by Board Shareownership



NOTE–These figures show how the effect of a dividend tax cut on dividends varies across firms with different ownership structures. In Figure 3a, the lower curve plots dividends versus the fraction of shares owned by the manager (α) when the tax rate is 40%. The upper curve plots the same when the tax rate is 20%. Figure 3b plots dividends versus the fraction of shares owned by the board of directors in the two tax regimes. Simulations are based on the same parametric assumptions as in Figure 2 along with $c(\gamma) = 10\gamma^2$.