An Agency Theory of Dividend Taxation*

Raj Chetty and Emmanuel Saez

UC Berkeley and NBER

October 14, 2007

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, chetty@econ.Berkeley.edu, Emmanuel Saez, saez@econ.berkeley.edu, University of California, Department of Economics, 549 Evans Hall #3880, Berkeley, CA 94720. 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.
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 the two leading existing 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.\textsuperscript{2}

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.\textsuperscript{3} 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.\textsuperscript{4} 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 would predict findings (3) and (4) on the cross-sectional heterogeneity in the dividend payout response by firm ownership structure.\textsuperscript{5}

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).\textsuperscript{6} Our model nests the neo-classical 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

\textsuperscript{2}Earlier work by Desai, Fritz, and Hines (2002) also found support for the agency view in the context of dividend remittances of foreign affiliates of U.S. multinational firms.

\textsuperscript{3}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.

\textsuperscript{4}Poterba's (2004) estimates using an old view model imply that the 2003 tax reform should increase dividend payments by 20 percent in the long run, but that the adjustment process is slow, with only a quarter of the long-run effect taking place within three years.

\textsuperscript{5}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.

\textsuperscript{6}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). 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.
shareholders does not matter for our analysis. Shareholders can provide incentives to managers 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 over-invest 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.\footnote{This result can be viewed as an application of 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 loss in surplus.}
However, if the contract between shareholders and the manager is not second best efficient, 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. 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. Second, Gordon and Dietz assume that dividend payout decisions are always made by shareholders (as opposed to management) to maximize their

---

8 An important limitation of the formula is that it does not account for share repurchases. Like most existing theories of dividend taxation, we abstract from share repurchase decisions in our model. See section 5 for a discussion of how share repurchases would affect the efficiency results.

9 We discuss the connections between our analysis and that of Gordon and Dietz in greater detail in sections 4 and 5.
total surplus. This assumption leads to very different results in both the positive and efficiency 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 \( r \).\(^{10}\)

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

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

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 \( D^* \):

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

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 \( IS \) 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

\(^{10}\)Throughout this paper, we abstract from general-equilibrium effects through which changes in \( t_d \) may affect the equilibrium rate of return, \( r \).
Auerbach (1979) and others. The source of this result is transparent in the two period case: 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

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

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

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:

\[ dW/dt_d = -P + P + t_d \frac{dP}{dt_d} = -\frac{t_d}{1 - t_d} \cdot \varepsilon_P \cdot P \] (5)

where \(\varepsilon_P = \frac{1 - t_d}{P} \frac{dP}{dt_d}\) 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 in 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 un-
dertaken 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.

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”.\footnote{There is a large literature in corporate finance providing evidence for such agency models. See e.g., Rajan et al. (2000), Scharfstein and Stein (2000), and Bertrand and Mullainathan (2003).} 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
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 $\alpha$ 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 $\gamma$ 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

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

(6)

Let $\omega = \alpha(1 - t_d) + \gamma$ denote the total weight that managers place on profits. When $\omega$ 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.\textsuperscript{12}

\textit{Shareholders’ Objectives.} Next, we characterize how shareholders choose the level of monitoring ($\gamma$). 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 $\alpha_i$ 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 $\gamma_i$ units of monitoring (the intensive margin) that satisfies $c'(\gamma_i = 0) = 0$. Increasing $\gamma$ is costly because reviewing managers’ plans, firing misbehaving

\textsuperscript{12}The pet project $g(J)$ is presumably small relative to the firm’s productive project $f(I)$. However, $\omega$ 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 $\omega$ can make the manager deviate substantially from the shareholders’ optimal investment level.
managers, etc. requires effort. Each shareholder chooses $\gamma_i$ to maximize his net profits

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

(7)

where $1(\gamma_i > 0)$ is an indicator function. In the Nash equilibrium, $\gamma$ is determined such that each shareholder’s choice of $\gamma_i$ 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. In addition, it is easy to see that there is a threshold level $\bar{\gamma}$ such that small shareholders with $\alpha_i < \bar{\gamma}$ will not monitor the firm, while large shareholders with $\alpha_i > \bar{\gamma}$ 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 $\gamma$ by forming a “board of directors” that is in charge of monitoring the manager. Let $\alpha_B$ denote the total fraction of shares held by the board of directors. The board chooses $\gamma$ 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:

$$V_B = (1 - t_d)\alpha_B \cdot [D(\omega) + \frac{1}{1 + r}f(I(\omega))] - c(\gamma)$$

(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 ($\alpha$ and $\alpha_B$) 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 board share ownership for the years in which data 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 \( \alpha, \alpha_B, \) 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 \( \alpha \) and \( \alpha_B \) as fixed in our positive analysis. In the longer run, and particularly when new firms are started, \( \alpha \) and \( \alpha_B \) are presumably endogenous to the tax regime. Therefore, in the efficiency analysis in section 5, we model how \( \alpha \) and \( \alpha_B \) are determined. Allowing for endogenous ownership structure is particularly important in the efficiency analysis because the deadweight cost of taxation depends critically on how \( \alpha \) and \( \alpha_B \) 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

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

(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:

\[
\omega f'(I) = g'(X - I - D)
\]  

(10)
\[ \omega \leq \frac{g'(X - I - D)}{1 + r} \text{ with strict equality iff } D > 0 \]  

(11)

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

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

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

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

**Proof.** Consider \( \omega \leq \underline{\omega} \). 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 \( \omega > \frac{g'(X - I^S)}{1 + r} = \underline{\omega} \), contradicting the supposition. Hence \( \omega \leq \underline{\omega} \Rightarrow D(\omega) = 0 \).

Now consider \( \omega > \underline{\omega} \). 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 \( \omega \leq \underline{\omega} \), contradicting the supposition. Hence \( \omega > \underline{\omega} \Rightarrow D(\omega) > 0 \), 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 \( \omega \) is below the threshold value \( \underline{\omega} \), 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 \( \omega \) 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 \( (\omega) \). Above \( \underline{\omega} \), increases in the weight on profits \( \omega \) lead to increases in dividends and reductions in pet investment on the intensive margin:

\[ D'(\omega) = -\frac{1 + r}{g''(J(\omega))} > 0 \text{ for } \omega > \underline{\omega} \]  

(12)

Now consider the manager’s investment choice. When \( \omega \leq \underline{\omega} \), 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 $\omega$ increases productive investment $I$ and reduces pet investment $J$:

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

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 ($\omega$) 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 $\alpha_B$ as fixed and chooses $\gamma$ to maximize

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

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

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

Intuitively, the board chooses $\gamma$ such that the marginal increase in the board’s share of profits by raising $\omega$ 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 \(\gamma\) is always in the interior, and hence (16) must be satisfied at the optimal level of monitoring \(\gamma(t_d)\). 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 \(\omega\), for any variable \(x \in \{ D, I, J \}\),

\[
\begin{align*}
\frac{dx}{dt_d} &= \frac{dx}{d\omega} \frac{d\omega}{dt_d}.
\end{align*}
\]

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 \(\frac{d\gamma}{dt_d}\), we implicitly difference the board’s first-order-condition for \(\gamma\) 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:

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

We know that the denominator of this expression is positive from the board’s second-order condition for \(\gamma\) in (16). The numerator is positive because \(P\) is increasing in \(\omega\) and \(c\) is convex. Equation (19) therefore shows that a reduction in the dividend tax rate leads to an increase in the weight \(\omega\) that managers put on profits. There are two channels through which this increase in \(\omega\) occurs. First, a decrease in \(t_d\) mechanically increases the net stake \((1 - t_d)\alpha\) that the manager has in the firm, effectively by reducing the government’s stake \((t_d)\) in the firm’s

\[\text{\footnote{The second order condition could hold with equality, a knife-edge case that we rule out by assumption.}}\]
profits. Second, a decrease in $t_d$ generally increases the level of monitoring $\gamma$ by the board.\footnote{Technically, it is possible to have $\frac{d\omega}{dt_d} > 0$ if the third derivatives $g''''(J), f''''(I), c''''(\gamma)$ are sufficiently large in magnitude. When these functions are quadratic, $\frac{d\omega}{\omega} > 0$ is unambiguously negative. Hence, barring sharp changes in the local curvature of the production functions, monitoring falls with the dividend tax rate.}

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 $\omega$, 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 $\omega \geq \overline{\omega}$. 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
three regions of $\omega$. 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 $\omega$ 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, the aggregate effect of a dividend tax cut on investment is ambiguous in the presence of agency problems. 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 $\omega(t_1) < \omega$, initiation likelihood increases with $\alpha$ and $\alpha_B$:

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

(ii) Extensive Margin: Size of Initiation. If $\omega(t_1) < \omega < \omega(t_2)$: $\frac{\partial \Delta D}{\partial \alpha} > 0$, $\frac{\partial \Delta D}{\partial \alpha_B} > 0$.

(iii) Intensive Margin. If $\overline{\omega} \leq \omega(t_1)$ and $g$ and $c$ are quadratic: $\frac{\partial D}{\partial \alpha} > 0$, $\frac{\partial D}{\partial \alpha_B} > 0$.

**Proof.**

(i) The result follows directly from the effect of $\alpha$ and $\alpha_B$ on $\omega$. 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 $\gamma$ in (16). Similarly,

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

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 $\alpha'$. Exploiting the result that $\frac{\partial \omega}{\partial \alpha_B} > 0$ yields the analogous result for $\alpha_B$.

(ii) When $\omega(t_1) < \omega < \omega(t_2)$, $D(t_1) = 0$ and hence $\Delta D = D(t_2)$. It follows that $\frac{\partial \Delta D}{\partial t_d} = \frac{\partial D(t_2)}{\partial t_d} = \frac{\partial D}{\partial t_d} \frac{\partial \omega}{\partial t_d}$ for $x \in \{\alpha, \alpha_B\}$. We know that $\frac{\partial D}{\partial t_d} > 0$ 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

\[
\frac{\partial D}{\partial t_d} = \frac{dD}{d\omega} \frac{d\omega}{dt_d} = \frac{1 + r}{g''(J(\omega))} \cdot \frac{\alpha \cdot c'' + \alpha_B P'(\omega)}{c'' - P'' \cdot \alpha_B(1 - t_d)}
\]
When \( \varpi < \omega \), \( P(\omega) = D(\omega) + \frac{f(\omega)}{1+r} \). 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 \( \alpha \) and \( \alpha_B \). Therefore, \( \frac{\partial^2 D}{\partial \alpha_d \partial \alpha} < 0 \) and \( \frac{\partial^2 D}{\partial \alpha_B \partial \alpha} < 0 \). Finally, \( \Delta D = (t_2 - t_1) \frac{\partial D}{\partial t_d} \) with \( t_2 - t_1 < 0 \), and therefore \( \frac{\partial \Delta D}{\partial \alpha_d} > 0 \) and \( \frac{\partial \Delta D}{\partial \alpha_B} > 0 \). QED.

Figure 3a plots \( D \) against \( \alpha \) 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 \( \alpha \)) are more likely to initiate dividend payments after the tax cut. This is because managers with higher \( \alpha \) are closer to the threshold (\( \varpi \)) of paying dividends to begin with, and are therefore more likely to cross that threshold. Second, conditional on initiating, firms with higher \( \alpha \) initiate larger dividends. Since \( D(t_2) \), the optimal dividend conditional on paying, is rising in \( \alpha \), the size of the dividend increase, \( \Delta D = D(t_2) \), is larger for firms with higher values of \( \alpha \) 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 \( \alpha \).\(^{15}\)

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 \( \omega \) when \( \alpha \) is large, the change in dividends is larger.

These three results apply analogously to the board’s shareholding (\( \alpha_B \)), as shown in Figure 3b. Non-paying firms with large \( \alpha_B \) are closer to the threshold \( \varpi \), 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 \( t_d \) thus has a greater effect on \( \gamma \) when \( \alpha_B \) is large, leading to a larger dividend response by the manager.

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

\(^{15}\)As above, this result holds as long as there are no sharp changes in the local curvature of the production functions. If the third derivatives \( g'''(J) \) and \( c'''(\gamma) \) are sufficiently large in magnitude, it is possible to have \( \frac{\partial^2 D}{\partial \alpha_d \partial \alpha_B} < 0 \).
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 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.\footnote{Firms with higher $X$ are closer to the threshold of paying dividends, for two reasons: (1) $\varpi$ is falling in $X$ and (2) $\gamma$ is rising in $X$. A tax cut is therefore more likely to make firms with higher $X$ cross the threshold and initiate dividend payments.} 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. In particular, 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 $\gamma$ and generates smaller $\Delta D$.\footnote{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.}
5 Efficiency Cost of Dividend Taxation

In this section we develop empirically implementable 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. As we discuss below, the efficiency results when α is fixed differ sharply from those when α varies endogenously with the tax rate. When both α and α_B are endogenous, it is convenient to write the formulas in terms of β_B = α_B / (1-α) – 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 (β_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 section. In the appendix, we show that two further extensions – corruption of the board and debt finance – do not affect this general formula.

5.1 Case I: Single External Shareholder [β_B = 1]:

Determination of α. We model the determination of managerial share ownership using the standard principal-agent model from the corporate finance literature. The manager’s stake α is chosen to maximize the firm’s value, taking into account the manager’s aversion to risk. 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 − π 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.

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

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

Let β = α(1 − t_d) denote the manager’s net of tax share of profits. As above, we assume
that the manager’s weight on profits is increased by shareholder monitoring \((\gamma)\), 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 \(\beta, \gamma, \text{ and } S\) 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 \(\beta, \gamma, \text{ and } S\).

The shareholder, who has linear utility, chooses \(\beta, \gamma, \text{ and } S\) to maximize his net payoff

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

subject to the participation constraint of the manager

\[
E u(D, I) = 0.
\]

The participation constraint pins down the value of \(S\) given a choice of \(\beta\) and \(\gamma\), 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 is to choose \(\beta\) and \(\gamma\) to maximize:

\[
W_S = (1 - t_d - \beta)P(\gamma, \beta) - c(\gamma)
\]

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 \((\alpha)\) while keeping expected the manager’s utility constant forces the principal to raise total expected compensation. Since raising compensation reduces net profits, the optimal \(\alpha\) 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 \( \beta \) and \( \gamma \) 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 that obtained from the old view model. As in the old view model, deadweight burden is a second-order function of the tax rate. That is, the marginal deadweight cost of taxation is typically small at low tax rates.

**Intuition: 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 and interesting 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, Auerbach and Hines, 2003, Goulder and Williams 2003). 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, because he does not fully internalize the shareholder’s welfare.

Why does our result differ from those of other studies in the tax literature? The pre-existing distortions analyzed in earlier studies are assumed to be exogenously fixed.
contrast, the pre-existing distortion in our model is endogenously determined through optimization of the manager’s contract to maximize net-of-tax surplus given the informational constraints. This endogenous determination of the distortion makes the deadweight cost of taxation second-order. To understand the intuition, observe that dividend taxation has a second-order efficiency cost as long as the following two conditions hold.

First, the tax should have no impact on contracting possibilities between the private agents. That is, the government should not be intrinsically better than the private market at resolving the informational constraints generating the distortion. Formally, shareholders must have a choice variable \( x \) such that \( k \frac{\partial W_S}{\partial x} = \frac{\partial W_S}{\partial t_d} + P \) where \( k \) is a constant. In the model above, the variable \( \alpha \) satisfies this condition. It is evident that the tax \( t_d \) has no effect on the contracting possibilities between the shareholder and the manager, because the manager’s decision program can be expressed as a function of \( \beta \) and \( \gamma \) (and not \( t_d \)). This eliminates all direct effects of \( t_d \) aside from the mechanical revenue effect from \( W_S \). Put differently, any effect of \( t_d \) on the contract beyond the mechanical revenue collection can be fully undone by the private market.

Second, the shareholder and manager must choose a contract that is second-best efficient, i.e. a contract that maximizes their joint surplus given the informational constraints. Formally, the requirement is that the variable \( x \) is chosen to maximize \( W_S \). To see why these two conditions imply a second-order efficiency cost, recall that

\[
\frac{dW}{dt_d} = \frac{dP}{dt_d} + P + \frac{dW_S}{dt_d} = \frac{dP}{dt_d} + k \frac{\partial W_S}{\partial x},
\]

where the latter equality uses condition 1. Next, note that condition 2 implies \( \frac{\partial W_S}{\partial x} = 0 \). Therefore \( \frac{dW}{dt_d} = t_d \frac{dP}{dt_d} \), which implies that the efficiency cost of dividend taxation is proportional to \( t_d^2 \). Intuitively, the government has no tools to affect the manager’s choice of dividends that the shareholders do not already have in our model. As long as the shareholders choose the manager’s contract to maximize private welfare, the only first-order effect of \( t_d \) on the shareholder’s utility 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 the government’s budget. This leaves only the term arising from the behavioral response of \( P \) to \( t_d \), which reduces government revenue. Since this term is proportional to \( t_d \), deadweight burden is second-order.

Traditional analyses of models with exogenous pre-existing distortions effectively assume
that the private sector has no tool to affect the size of the distortion. In that setting, the
government has a technological advantage relative to the private sector, and can introduce first-
order changes in efficiency through taxes and subsidies. The same applies in our analysis: if
\( \alpha \) were exogenously fixed at some level \( \alpha_0 \) 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 \( \alpha \) 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 \( \gamma \) 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 adequate to infer that government taxes or subsidies will have
first-order effects on the economy, contrary to conventional wisdom. It is critical to understand
the private sector’s ability to affect the size of the distortion – specifically whether the tools
available to the private sector and government are the same and whether the private sector
reaches the second-best efficient outcome.\(^{18}\)

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. Hence, all the
models we will consider satisfy the first condition.\(^{19}\) However, it is quite likely that the
second condition – private contracts maximizing total net-of-tax surplus – breaks 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 Case II. Dispersed Shareholders \([\beta_B < 1]\)

We now consider the case where the fraction of shares owned by the board \((\beta_B)\) is fixed at a
level less than 1, so that there are some dispersed “minority” shareholders who do not monitor

\(^{18}\)This result can be connected to the analysis of the first welfare theorem in economies with private informa-
tion (Prescott and Townsend, 1984a, b). When such a constrained first welfare theorem holds, a price distortion
created by a small tax generates a second order effect on welfare.

\(^{19}\)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.
the firm in equilibrium. The manager’s objective remains the same as above for given values of \( \beta = \alpha(1-t_d) \) and \( \gamma \). The board of directors chooses \( \beta \) and \( \gamma \) to maximize its own welfare:

\[
W_B = \beta B(1 - t_d - \beta)P(\gamma, \beta) - c(\gamma)
\]

This objective differs from the objective used to choose \( \beta \) and \( \gamma \) in the single shareholder model above 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 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:

\[
W = t_dP + W_B + W_M
\]

It is easy to see that \( dW_B/dt_d = -\beta_B P \) because of the envelope conditions for \( \beta \) and \( \gamma \) from board maximization. Furthermore, because \( \beta \) maximizes \((1 - t_d - \beta)P(\gamma, \beta)\), we have:

\[
\frac{dW_M}{dt_d} = -(1 - \beta_B)P + (1 - \beta_B)(1 - t_d - \beta)\frac{dP}{d\gamma}\frac{d\gamma}{dt_d}.
\]

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

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

Equation (24) shows that when shareholders are dispersed, the deadweight cost of raising the dividend tax is the sum of the second-order term obtained with a single owner and a new term proportional to the fraction of dispersed shareholders \((1 - \beta_B)\). Since this new term is not proportional to \( t_d \), it leads to a first-order deadweight cost of dividend taxation. The deadweight burden is first-order when \( \beta_B < 1 \) because the board chooses \( \gamma \) to maximize \( W_B \) rather than total shareholder surplus \((W_B + W_M)\). As a result, the manager’s contract is no longer second-best efficient, breaking the second condition described in the previous section.
To understand the intuition for the first-order term more concretely, note that the board picks \( \gamma \) without taking account of the spillover benefits of monitoring to the dispersed shareholders. Since the dispersed shareholders prefer to free-ride, the equilibrium outcome is a level of monitoring that is suboptimal even relative to the second-best outcome where total social surplus is maximized net of monitoring costs. This 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. A dividend tax moves precisely in the opposite direction by further reducing the incentive to monitor, and thus creates a first-order efficiency cost.

Note that in the first-order term in (24), the derivative of \( P \) with respect to \( t_d \) is taken keeping \( \beta \) constant. This is because the level of \( \beta \) chosen by the board is optimal from the minority shareholders’ perspective as well, since their payoffs are proportional to each other aside from monitoring costs. The level of \( \beta \) 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 \( \alpha \) dilutes the shareholding of both small and large shareholders proportionally. In contrast, the board is not able to share the cost \( c(\gamma) \) with the minority shareholders, and therefore sets \( \gamma \) at a suboptimally low level. Thus, only the effect of \( t_d \) on \( P \) through \( \gamma \) leads to an externality effect. This result confirms the critical role of second-best inefficiency of contracts in generating a first-order efficiency cost of taxation.

Gordon and Dietz (2006) assume that the board of directors acts on behalf of all shareholders. Therefore, in their model, there is no free riding monitoring issue and hence they do not obtain a first order efficiency cost of taxation as we do here.

An Alternative Representation. In order to implement empirically (24), 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 \( \beta \) in the board’s objective implies that \(-P + (1 - \beta \sum \alpha \frac{d \alpha}{dt_d} ) = 0\).
Recognizing that $\beta = \alpha (1 - t_d)$, it follows that

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

(25)

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. Using the elasticity of dividend payments with respect to the net-of-tax rate as $\varepsilon_P,1-t_d$, we can rewrite (24) as

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

(26)

Equation (26) is easier to implement empirically than (24) because one could in principle estimate how managerial shareownership ($\alpha$) responds to changes in $t_d$ to calculate $\varepsilon_{\alpha,1-t_d}$.

5.3 Case III. Equity Issues and Endogenous $\beta_B$

To complete our analysis, we now consider a model in which the ownership structure of the firm is fully endogenous, i.e. both $\alpha_B$ and $\alpha$ are determined endogenously. When $\alpha_B$ 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 $\beta_B$, and then turn to the efficiency analysis. The key result is that allowing for endogenous $\alpha_B$ and $E$ does not affect the formula for deadweight burden in (24).

Model of $E$ and $\beta_B$. 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 Cases I and II analyzed above. In the first stage, the external shareholders choose $E$, and an acquirer buys a fraction $\beta_B$ of the outstanding shares to take control of the board and set the manager’s contract ($\beta, \gamma$). 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 $\beta_B$ of the company and making a contract ($\beta, \gamma$) with
the manager. The acquirer picks \((\beta_B, \beta, \gamma)\) 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 works backwards. Conditional on \(\beta, \gamma,\) 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. The 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 the amount of equity raised.

To model how \(\beta_B\) 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.\(^{20}\)

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 \(\beta_B\) of the company \(K(\beta_B)\). The function \(K\) reflects the value of controlling a large company, e.g. through perks or directly utility of control.\(^{21}\) Presumably, \(K(\beta_B)\) is first increasing in \(\beta_B\) — as a very low value of \(\beta_B\) would not yield any control — and then decreasing once \(\beta_B\) is large, since complete control can be achieved with less than a 100% stake and liquidity constraints or lack of diversification could make a large \(\beta_B\) costly to a risk averse shareholder. This inverted-U shape of \(K\) leads to an interior optimum in \(\beta_B\). Taking \(E\) and \(t_d\) as given, the board of directors chooses \(\beta_B, \beta\) and \(\gamma\) to maximize

\[
\beta_B[(1 - t_d - \beta)P(\gamma, \beta, E) - P_*] - c(\gamma) - K(\beta_B) \tag{27}
\]

where \(P_*\) is the equilibrium price of the firm, which the acquirer takes as fixed when making

\(^{20}\)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).

\(^{21}\)In the appendix, we show that modelling such benefits as kickbacks from the manager to the large shareholder (a “corrupt board”) produces the same results.
the tender offer to the dispersed shareholders. The first order conditions with respect to $\beta$, $\gamma$, and $\beta_B$ are:

\[
(1 - t_d - \beta) \frac{\partial P}{\partial \beta} = P
\]

(28)

\[
\beta_B(1 - t_d - \beta) \frac{\partial P}{\partial \gamma} = c'(\gamma)
\]

(29)

\[
(1 - t_d - \beta)P(\gamma, \beta, E) - P_* = K'(\beta_B)
\]

(30)

In equilibrium, $P_* = (1 - t_d - \beta)P(\gamma, \beta, E)$. Hence, the first order condition for $\beta_B$ simplifies to $K'(\beta_B) = 0$, which implies that $\beta_B$ 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 $\beta_B$.

Finally, the dispersed shareholders choose $E$ to maximize their total dividend payout net of equity investment. We assume that dispersed shareholders anticipate the takeover by the large shareholder when choosing $E$ but do not internalize the effect of $E$ on $\beta$ and $\gamma$ because they are small. Hence, $E$ is chosen to maximize

\[
W_M = (1 - t_d - \beta)P(\beta, \gamma, E) - E
\]

(31)

The first order condition with respect to $E$ is:

\[
(1 - t_d - \beta) \frac{\partial P}{\partial E} = 1,
\]

(32)

The solution to (32) yields a function $E(t_d, \beta, \gamma)$. The Nash equilibrium levels of $\beta$, $\gamma$, and $E$ are given by a triplet $(\beta_*, \gamma_*, E_*)$ such that each agent’s behavior is optimal given the other’s choice:

\[
\beta_* = \beta(t_d, E_*), \quad \gamma_* = \gamma(t_d, E_*), \quad E_* = E(t_d, \beta_*, \gamma_*).
\]

The equilibrium triplet is a function of $t_d$, which is 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

\[ W = t_d P + V(t_d) - E - K(\beta_B) - c(\gamma) \]  

(33)

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

**Proposition 3** The excess burden with endogenous ownership structure and equity issues is given by

\[ \frac{dW}{dt_d} = t_d \frac{dP}{dt_d} + (1 - \beta_B)(1 - \alpha)(1 - t_d) \frac{dP}{dt_d} \bigg|_{\beta, E}. \]  

(34)

**Proof.** Totally differentiating \( V \) with respect to \( t_d \) gives

\[ V'(t_d) = -P + \left[ -P + (1 - t_d - \beta) \frac{\partial P}{\partial \beta} \right] \beta^*(t_d) + (1 - t_d - \beta) \frac{\partial P}{\partial E} E^*(t_d). \]

Using the first order conditions (28) and (32), this expression simplifies to:

\[ V'(t_d) = -P + (1 - t_d - \beta) \frac{\partial P}{\partial E} E^*(t_d). \]

Next, using (29), differentiating (33) implies that

\[ \frac{dW}{dt_d} = t_d \frac{dP}{dt_d} + V'(t_d) - E^*(t_d) - c'(\gamma) \gamma^*(t_d) \]

(34)

which can be rewritten as (34). QED.

Equation (34) is exactly the same formula that we obtained in equation (24) with fixed \( \beta_B \) and no equity issues. Intuitively, introducing equity issues does not affect the formula for \( \frac{dW}{dt_d} \) because there is no externality in the equity issue decision in our model. That is, \( E \) is set at the second-best efficient level that maximizes total net-of-tax surplus. This is because shareholders, and not entrenched managers, decide how much equity is given to the firm. 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 also explains why the first-order
term in (34) depends on $dP/dt_d$ keeping both $\beta$ and $E$ fixed, as only the $\gamma$ channel is subject to the free-rider inefficiency problem.\footnote{This value of $dP/dt_d|_{\beta,E}$ can be estimated empirically using the same method as in Case II. 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 condition (32), just as we removed the $\beta$ channel in (26).}

Allowing for endogenous determination of $\beta_B$ does not affect the formula because the acquirer chooses $\beta_B$ 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 distorts the choice of $\beta_B$ and $\gamma$, both of which are taken into account in the empirical estimate of $\frac{dP}{dt_d}|_{\beta,E}$, leaving the formula for $dW/dt_d$ unchanged.

It is useful to contrast this result with the effect of allowing for endogeneity of $\alpha$. In case I, we showed that deadweight burden is first-order when $\alpha$ is fixed but second-order when $\alpha$ is endogenous. Allowing for endogenous determination of $\alpha$ does affect the formula because $\alpha$ is chosen to maximize social surplus in our model. In a model with multiple external shareholders, the costs of raising $\alpha$ are shared through dilution, and thus $\alpha$ is set at the second-best efficient level. In contrast, standard economic principles suggest that $\beta_B$ will not be chosen to maximize social surplus, since the costs of controlling and monitoring a firm are borne only by the large shareholder. This contrast in the results on $\alpha$ and $\beta_B$ underscores the central role of the second-best efficiency of private contracts in determining the efficiency cost of taxation in a market with pre-existing distortions.

5.4 Extensions

In the appendix, we analyze two additional extensions of the model to evaluate the robustness of the formula in Proposition 3. First, we allow the firm to raise funds through both debt ($L$) and equity. 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 (Jensen and Meckling, 1976). We model the cost of carrying debt in reduced-form through a convex cost function $c_L(L)$. As in the model above, the dispersed shareholders choose both $E$ and $L$ in the first stage; all other stages remain identical.

Second, we 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 strictly. We model such corruption by adding a term $\sigma g(J)$ to the board of director’s welfare in (27).

In both of these extensions, we find that the formula for $\frac{dW}{dt_d}$ is unchanged, though the values of the inputs (e.g. $\varepsilon_{P1-L_d}$) will generally differ. Since these inputs are estimated empirically rather than simulated from assumptions about deep structural parameters, our method of computing the efficiency cost of taxation is robust to these variations. The formula is robust because the envelope conditions used to derive (34) are unaffected by introducing additional choice variables or constraints into the manager’s and shareholders’ problems, as long as these choices do not have externalities on other agents.

Following standard practice in the public finance literature, we have ruled out 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 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 (34) 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.

Since there is no widely accepted theory that explains why companies pay dividends despite the tax advantage of repurchases, we do not attempt to characterize the efficiency cost of dividend taxation in an environment with repurchases here. In future work, it would be useful to understand the costs and benefits of share repurchases and incorporate them into an agency model to further improve our understanding of the efficiency costs of dividend taxation.
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: (1) $\varepsilon_{P,1-t_d}$ – the long-run elasticity of dividends with respect to the net-of-tax rate, (2) $\alpha_B$ – the fraction of shares owned by those on the board of directors, (3) $\alpha$ – the fraction of shares owned by the manager, and (4) $\frac{d\alpha}{dt_d}, \frac{dE}{dt_d}$ – the effects of $t_d$ on $\alpha$ and $E$. The remaining structure of the model – e.g. the monitoring and share acquisition cost functions $c(\gamma)$ and $K(\beta_B)$, the pet project payoff $g(J)$, or the level of monitoring $\gamma$ – are irrelevant conditional on knowing these values. This is important because estimating these functions would be difficult, particularly since they represent reduced forms of complex contracts and payoffs for shareholders, the board of directors, 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 ignore equity issues in this calculation. Note that managerial share ownership 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.\footnote{Although our calculation focused solely on stock ownership, accounting for other forms of incentive-based pay is unlikely to raise $\alpha$ significantly. Existing studies have measured $\alpha$ more broadly by computing the change in the wealth of CEOs when the firm’s value increases by $1$. These studies estimate that $\alpha$ is less than 1% on average for CEOs of publicly traded corporations in the U.S. (see Murphy 1999 for a survey).} Therefore, as an approximation we assume $\alpha = 0$. This simplifies the formula in (26) to:

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

The marginal revenue from raising the dividend tax rate is

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

Using data on the shareholding of members of the board of directors and other large (5%+) blockholders of companies from Dlugosz et. al. (2004), we estimate that $\alpha_B = 0.1$ on average between 1998 and 2005 (see data appendix for details). Plugging $\alpha_B = 0.1$, $\varepsilon_{P,1-t_d} = 0.75$ (Chetty and Saez 2005), and $t_d = 0.15$ into (35) and (36), 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. The 2003 dividend tax cut cut the tax rate on qualified dividends paid to individuals by about 20 percentage points. Qualified dividend payments to individuals who benefited from the tax rate reduction totalled \( P = \$119 \) billion in 2005 (US Treasury Department: Internal Revenue Service, 2007). Using the linear approximation of \( W(t_d) \) in (35), 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. We estimate that this tax increase would produce approximately \( \$21 \) billion of additional government revenue in the short run.

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, and particularly when new firms are started, \( \alpha \) 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 entirely signalling issues. 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 social welfare perspective.

\(^{24}\)More work is needed to integrate the signalling and agency theories of dividend taxation as both theories have received empirical support.
In view of these caveats, we believe that this analysis is most useful in providing two qualitative lessons. First, the main source of inefficiency from taxing dividends is the under-monitoring of managers because of the free-rider problem, and not the classic distortion to revenue 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 in raising revenue 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 corrective dividend subsidy may be desirable to subsidize monitoring and correct the free-rider problem.

7 Conclusion

The public finance literature on corporate taxation is based primarily on neoclassical 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. The broad aim of this paper has been to take a step toward bridging this gap. We analyzed the effects of dividend taxation in a model in which managers’ interests differ from shareholders. The agency model generates a set of predictions that explain several pieces of empirical evidence that pose problems for neoclassical models.

The agency theory yields new insights into the efficiency costs of dividend taxation. The main result is that the efficiency consequences of dividend taxation depend 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. In this environment, dividend taxation has a first-order efficiency cost because it exacerbates the under-monitoring problem.

Our analysis suggests that agency issues may be much more important factors in the excess burden of taxation than Harberger revenue distortions. Further research on the interaction
between taxes and agency problems could yield new insights into the welfare consequences of other important corporate tax policies, such as the corporate tax, the capital gains tax, and tax treatment of interest payments.
Appendix A: Extensions of Efficiency Analysis

Corrupt Board.

Consider the model in Case II with fixed $\beta_B$ for simplicity; as above, allowing for endogenous determination of $\beta_B$ 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 $\sigma g(J)$ into the board’s welfare. This changes the board’s objective when setting $\alpha$ and $\gamma$ 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 the basic model conditional on $\beta$ and $\gamma$. Total surplus is $t_dP + W_B + W_M$. As is now familiar, we exploit the envelope conditions for $\gamma$, and $\beta$ 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 (26). The $\sigma g(J(\gamma, \beta))$ term can effectively be viewed as a modification of the cost function $c(\gamma)$. Since this function does not enter our formula for deadweight burden, allowing for internal corruption via $\sigma > 0$ does not affect the results.\(^{25}\)

Debt Finance.

There is a large literature in corporate finance on the capital structure of firms (see e.g., Jensen and Meckling 1976 or more recently Zwiebel, 1996). In this section, we assume that the firm can raise additional funds for investment from the external capital market. The firm raises 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. Following the standard explanation of why firms do not finance all investment using debt, we assume that issuing debt has costs (e.g. risk of bankruptcy), which we model through a reduced-form function $c_L(L)$. The firm’s total cash available for dividends and investment is $X + E + L$

\(^{25}\)Note however, that formula (24) with $dP/dt_d|_{\beta}$ is no longer valid because the corruption effect $\sigma g(J(\beta, \gamma))$ introduces a negative externality in the choice of $\beta$ as well which has to be taken into account in the marginal deadweight burden computation.
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 and shareholder’s choices can be broken into two stages as in the Case III studied in the text. First, there is a Nash 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 \((\beta_B)\) and chooses a contract \(\alpha, \gamma\) with the manager. Second, the manager chooses \(D\) and \(I\).

Conditional on \(\beta, \gamma, E, L\) 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 Case III in the text, conditional on \(E, L\), and \(t_d\) the board of directors chooses \(\beta, \gamma, \) and \(\beta_B\) 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 \(\beta_B\) is again \(K'(\beta_B) = 0\) which shows that \(\beta_B\) is independent from other variables.

Finally, the dispersed shareholders choose \(E\) and \(L\) to maximize their total payout net of equity investment conditional on \(\beta, \gamma, \) and \(t_d\):

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

Those maximization programs lead to a Nash equilibrium \((\beta_*, \gamma_*, E_*, L_*)\) which depends on \(t_d\). We denote again the net value of the firm \((1 - t_d - \beta)P(\beta, \gamma, E, L)\) in equilibrium by \(V(t_d)\).

Social surplus is defined as \(W = t_dP + W_B + W_M = t_dP(t_d) + V(t_d) - E_*(t_d) - c(\gamma_*(t_d)) - K(\beta_B)\) as in Case III above. Because, \(L\) maximizes \(P(\beta, \gamma, E, L)\) so that \(\partial P/\partial L = 0\), the same computation as in case III can be repeated to evaluate \(V'(t_d)\) and \(dW/dt_d\) and we obtain exactly the same expression as (34).

**Appendix B: Data for Figure 1 and Table 2**

39
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 weights equal to the average end-of-year 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 (accessed 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 the effect of data errors, a small number of observations are omitted if the 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 end-of-year market capitalization between 1996 and 2005.

Dividend Payments. Data on dividend payments comes 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 $\alpha_B$ in Section 6. The value for $\alpha_B$ 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. (2004), available through WRDS.

The average level of non-employee director ownership between 1998 and 2005 is 2.4%. The average level of outside blockholder ownership (as defined by the variable SUMOUT) between 1998 and 2001 (the last year of data) 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


Korinek, Anton and Joseph Stiglitz (2006), “Dividend Taxation and Intertemporal Tax Arbi-

42


### Table 1
Summary of Key Predictions: Neoclassical vs. Agency Models

<table>
<thead>
<tr>
<th>Initial Cash X</th>
<th>Neoclassical Model</th>
<th>Agency Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Old View</td>
<td>New View</td>
</tr>
<tr>
<td>Low: $f'(X) &gt; (1+r)/(1-t_d)$</td>
<td>High: $1+r&gt;f'(X)$</td>
<td></td>
</tr>
<tr>
<td>Medium: $(1+r)/(1-t_d) &gt;= f'(X) &gt;= 1+r$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High: $1+r&gt;f'(X)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Dividends D | $D=0$ | $D>0$, $f'(X-D)=1+r$ | $D=0$ | $D>0$, $g'(X-I^S-D) = \omega(1+r)$ |
| Equity Issues E | $E=0$, $f'(X+E) = (1+r)/(1-t_d)$ | $E=0$ | $E=0$ |
| Investment I | $I=X+D>X$, $f'(I)=f'(X)$ | $I=X-D<X$, $f'(I)=1+r$ | $I=X-D<X$, $f'(I)=1+r$ |

<table>
<thead>
<tr>
<th>Effects of reducing dividend tax $t_d$</th>
<th>Neoclassical Model</th>
<th>Agency Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>I increases, E increases</td>
<td>No effect on D</td>
<td>Extensive margin: No effect on D and E, I increases, J decreases</td>
</tr>
<tr>
<td>I increases, E increases</td>
<td>No effect on D, E, I</td>
<td>Extensive margin: Some firms shift to very high cash regime, start paying dividends</td>
</tr>
<tr>
<td>Efficiency cost of $t_d$</td>
<td>Harberger</td>
<td>Harberger (along extensive margin)</td>
</tr>
</tbody>
</table>

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 cost $t_d$ 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

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log Executive Ownership Share</td>
<td>Log Director Ownership Share</td>
<td>Log Dividend Payments</td>
</tr>
<tr>
<td>Post-2003 Indicator</td>
<td>0.048 (0.063)</td>
<td>-0.048 (0.075)</td>
<td>0.148 (0.043)</td>
</tr>
<tr>
<td>Year</td>
<td>-0.083 (0.010)</td>
<td>-0.014 (0.012)</td>
<td>-0.014 (0.043)</td>
</tr>
<tr>
<td>Constant</td>
<td>166.061 (20.129)</td>
<td>29.367 (24.173)</td>
<td>40.246 (13.897)</td>
</tr>
</tbody>
</table>

### 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.
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, \( \omega \). 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 \( r = 0 \).
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 ($\alpha$) 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 $c(\gamma) = 10\gamma^2$. 