A MODEL OF MARKET DISCIPLINE

Colin Ward and Chao Ying*

University of Minnesota

22 October 2020

Abstract

We develop an equilibrium model where cash holdings, costly refinancing policies, and managerial incentives are jointly determined to quantify the market’s influence on management’s ex ante behavior. We also derive a general formula that shows how agency and financing distortions shape payouts and compensation, two easily measured quantities. Our calibrated model estimates agency conflicts are nearly 10 times more severe than financial frictions for US public firms. Our analysis suggests that cutting corporate income taxes while introducing a tax on refinancing can reduce the relative severity of agency.

*Contact address: Colin Ward, Carlson School of Management 3-127, 321 19th Avenue South, Minneapolis, Minnesota 55455; email: cward@umn.edu; phone: (267) 242-4834. We thank Hengjie Ai, Yasser Boualam, Barney Hartman-Glaser, Burton Hollifield, Martin Szydlowski and participants at Minnesota (Carlson) and SFS Cavalcade for useful suggestions.
INTRODUCTION

Crucial to the goal of corporate value maximization is the disciplinary role of markets on aligning investors’ and managers’ incentives. Even though infrequently tapped for capital, markets invisibly guide management’s use of resources, as serially poor judgment eventually necessitates costly refinance. The degree to which this channel operates, however, is an open question. How strong is it and how does it manifest itself? More broadly, how do agency and financial frictions jointly affect firm behavior and can they be altered? In spite of the importance of these questions concerning the efficacy of markets, a framework suitable for studying them has proved challenging.

In this paper, we attempt to fill this gap. We do so by developing a quantitative model that combines a dynamic agency problem with internal and costly external finance. In the model, investors anticipate refinancing’s effect on management’s incentives and thus judiciously choose it given the firm’s history. Managers understand this and, therefore, it affects their behavior today and subsequent real outcomes, which in turn feeds back into investors’ expectations. We call this equilibrium effect market discipline.

Our paper makes two major contributions. First, by formalizing the interaction between financial and agency frictions, we provide a unified framework for assessing the role of markets in shaping cash holdings, investment, payouts, compensation, and whether to refinance a firm or let it fail. Second, we derive a general formula that connects the relative sizes of agency conflicts and financial frictions to the relative allocation of free cash flow across investors and managers:

\[
\frac{\text{Size of Agency Conflict}}{\text{Size of Financial Friction}} \propto \frac{\text{Marginal Cost of Delaying Payouts to Investors}}{\text{Marginal Cost of Delaying Payments to Managers}}.
\]

For intuition, consider a firm with scarce cash holdings. Here the marginal cost of delaying payouts to investors is low as the likelihood of costly refinancing is large, and preferably avoided altogether. Any payments made to managers, during which agency conflicts are muted, must then correspond to a financial distortion. The idea is akin to using a supply shock to identify a demand elasticity and analogously implies that agency conflicts can be inferred using payouts. Thus, we show that the scale at which managers or investors are differentially paid are informative about the relative magnitude of these underlying frictions.

It is useful at this point to take a step back into the model setup of Section I to better understand our headline results. To separate interests and motivate their desire to grow the firm beyond the optimal size, we allow managers the possibility of consuming private benefits that are increasing in firm assets and resources (Jensen (1986)). Their private consumption comes at the expense of their effort to increase asset efficiency and enhance the probability of firm survival. Investors therefore

---

1If only costly financing were specified, markets would have no scope to affect management’s already optimal behavior. If only an agency conflict were present, instances of costly refinancing would be left indeterminate and remain non-measurable. The ability to finance internally makes this costly event meaningful.
write a contract that continually provides managers with the incentive to choose the appropriate action.

The contract is history-dependent and acts as the bridge through which managers internalize investors forward-looking rational expectations. The dynamics of managerial incentives, in turn, influence several firm policies including refinancing, and not just the nearest refinancing event but over all those that potentially could occur afterwards.

Our model’s solution described in Section II takes the form of a partial differential equation summarized by two state variables, each normalized by assets: cash holdings and managers’ stake, which measures their effective ownership of the firm and proxies for both compensation and the (inverse) agency problem in this dynamic environment. The model endogenizes four decision thresholds that collectively provide lower and upper bounds for each of these states: (1) when investors collect payouts out of cash versus (2) when the firm is refinanced; and (3) when management receives payments versus (4) when the firm is liquidated. A key challenge that we overcome is thus determining the value of the solution jointly with the shape of the state space.

Firms strive to reach the state where positive free cash flow is freely paid to both investors and managers; that is, the upper thresholds (1) and (3) apply. At this joint upper boundary, the general formula holds explicitly and with equality in our setup. It equates the allocation of free cash flow across players to the ratio of two lingering marginal costs: a tax penalty to holding cash in the firm and managers’ relative impatience to investors. This ratio also defines the linear span of a second-best frontier that bounds from above and is tangent to the endogenous state space. Novelly, we quantify the size of agency conflicts and financial frictions as proportional to the distance between the upper thresholds and the frontier.

Next, moving away from the joint upper boundary (where (1) and (3) hold) to along one upper boundary (where (1) or (3) holds) implies one distortion will remain minimal while the other is aggravated. The boundary’s shape is therefore informative on the relative frictions that firms currently face and naturally extends to relative allocation of free cash flow, thus closing the step in understanding the origins of the general formula.

Our measurement approach has advantages over those obtainable in classic agency or financial friction models. First, classic models measure frictions as deviations from first-best, which may be neither attainable nor serve as a reasonable benchmark. Second, they often make comparisons implicitly based on market values that are influenced by hard-to-measure discount rates. By contrast, we measure relative to the second-best frontier and we do so with quantities of cash flows that are readily observable. That payouts and payments can be observed with simple accounting data is an appealing feature with empirical potential, but is not within this paper’s scope to pursue.

More broadly, a burgeoning literature studies the impact of financing frictions on the economy. It is unclear that in developed economies, however, that these are the utmost concern since capital
markets are quite deep. Instead, we argue that agency frictions are likely to be more onerous. Indeed, our calibrated model of Section III suggests that they are nearly 10 times more severe than financial frictions in the United States. This is because while cash can simply be accumulated to minimize financing frictions, it is double-edged as it exacerbates the alignment of managerial incentives.

Part of our calibration is done internal to the model and targets moments in the data that speak to the rich features of market discipline—payouts to investors, managerial compensation, and frequencies of refinancing and liquidation, among others—outcomes that are obviously important to corporate finance and firm value maximization but have been little, if at all, studied together in modern structural models.

We then proceed to evaluating model counterfactuals in Section IV by conducting steady state analysis in the spirit of Hopenhayn (1992) by examining how the stationary distribution shifts in response to a change in parameter. The stationary distribution encodes all of the information about the stochastic environment and policy functions of the model solution and is therefore an ideal object to study.

Among other analysis, we examine a policy counterfactual where we lower the corporate tax rate from 30 to 21 percent and raise the external cost of finance from 50 to 150 basis points, which could be implemented with a small tax on the event of refinancing. In effect, lower corporate taxes help offset the cost that investors would otherwise bear for a refinancing tax and further allows them to allocate additional cash flow to mitigating agency conflicts. While subject to caveats that we discuss in the paper, we find that its stationary equilibrium mimics an economy where relative agency frictions are reduced by third.

**Literature**

Our paper analyzes the capital market implications of dynamic agency. As in DeMarzo, Fishman, He and Wang (2012), we analyze investment in the context of a dynamic agency model. In their paper, the optimal contract relaxes agents’ incentive constraint following a history of good shocks, which raises the marginal benefit of investing in more capital. However, in this paper the distinction between internal and external sources of finance are left unexplored, therefore ignoring the ex ante effect that discrete instances of refinancing have on agents’ incentives.

Zwiebel (1996) critiques that a recurring, and problematic, feature of traditional agency models is that a “discipliner” is present ex ante yet absent ex post. Often in these models the discipliner sets constraints (for example, debt) that ex ante restrict managers’ future decisions. If instead
the discipliner were present ex post, management could still be restricted even though constraints were never set. He argues that the correct formulation of constraints, whether ex ante or ex post, is dynamically consistent, as they are in our model. Our contribution here is that our paper studies a broader range of corporate policies in the context of quantitative model.

Hartman-Glaser, Mayer and Milbradt (2019) study moral hazard’s effect in an environment where the firm accumulates cash, similar to ours. They show that when cash holdings are low, firms transfer cash flow risk to managers, hoping to minimize their desire to divert cash. In addition, they show that permitting the payments of small, negative wages to managers allows them to solve the model as a function of only one state variable. In contrast, we solve the model with two states without resorting to a restricted problem. Another novelty of our paper is that the event of refinancing itself is endogenous, a key feature in isolating the market’s disciplinary effect on management’s behavior.

Our theory complements the literature on financing constraints. Bolton, Chen and Wang (2011) show that the marginal value of cash affects investment, external financing, and risk management. Specifically they show that cash holdings follow a fixed double barrier policy. The lower bound has the firm either refinancing or liquidating, depending on the choice of parameters. In contrast, our double barriers are dependent on the level of management’s incentives, allowing us to study the interaction between financial and agency frictions. In addition, the important decision whether to let the firm refinance or fail is endogenous to our model.

Our paper contributes to the literature on misallocation and the measurement of distortions. Pioneered by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), a large part of the literature has focused its determinants such as financial frictions (for example, Moll (2014), Midrigan and Xu (2014), and Gopinath, Kalemlı-Ozcan, Karabarbounis and Villegas-Sanchez (2017)). Our model, in contrast, jointly determines financial frictions with agency conflicts, a feature largely ignored in the literature. We thus provide a new rationale that possibly contributes to the mysterious decline in US allocative efficiency (Bils, Klenow and Ruane (2020)).

I. MODEL

Here we present the model’s setup. We first describe the firm’s technology and how managerial effort affects its efficiency. We then introduce the flow equation for resources and costs of external financing. We finally discuss the agency problem and close the setup within the contracting

---


environment.

A. TECHNOLOGY, FREE CASH FLOW, AND MANAGEMENT’S EFFORT

Capital, $K$, is used to produce output and evolves according to the standard accumulation equation

$$dK_t = (I_t - \delta K_t) dt, \tag{1}$$

where $I$ is gross investment and $\delta \geq 0$ is the rate of depreciation. Following the literature on $q$ theory (Hayashi (1982) and Abel and Eberly (1994)), investment incurs adjustment costs $G(I, K)$, allowing us to write the total cost of investment as $I + G(I, K)$.

Free cash flow is the cash flow available to distribute to investors after paying taxes at rate $\tau_Y$ and paying for the investment and expenses required to maintain the firm’s existing operations. After optimally choosing and compensating freely adjustable labor (non-management), free cash flow is determined by a constant returns to scale technology:

$$dY_t = (1 - \tau_Y)dA_t K_t - I_t dt - G(I_t, K_t) dt. \tag{2}$$

Asset productivity is determined by management’s unobservable effort, $e_t \in \{0, 1\}$:

$$dA_t = e_t \mu dt + \sigma dZ_t, \tag{3}$$

where $dA$ is a productivity shock with drift $e_t \mu \geq 0$ that varies with a standard Brownian increment $dZ$ scaled by volatility $\sigma > 0$. Effort ($e_t = 1$) enhances profits and the likelihood of firm survival. However, because only $dA$ will be observable and contractible to investors, the extent to which management can hide their effort choice and shirk ($e_t = 0$) will scale with $\sigma$.

Even though the economic environment is $iid$, policies dictating the firm’s experience under the optimal contract will be path dependent. Notably, the firm’s cash flow history, which is in part determined by managerial effort, will affect equilibrium outcomes such as refinancing.

B. INTERNAL RESOURCES, DISTRIBUTIONS, AND COSTLY REFINANCING

Donaldson (1984) describes a firm’s resources as “the aggregate purchasing power available to management for strategic purposes during any given planning period.” They are thus not limited to pecuniary things and could reflect the sophistication of business networks or even the operational

---

5For a given capital stock and with freely adjustable labor, $L$, the firm solves the static problem $\max_L (1 - \tau_Y)(a_t K_t^\alpha L_t^{1-\alpha} - w_t L_t)$, where $a_t$ is a productivity shock and $w_L$ is the wage rate which could be stochastic. The optimal labor choice will be proportional to capital. The productivity shock $dA_t$ used elsewhere thus depends on $a_t$, $w_L$, and $\alpha$. 

---
efficiency of the firm—any of these could be squandered by managers. To empirically construct a law of motion, however, we restrict them to be only cash holdings.

Cash held at time $t$ is denoted by $C_t$ and the flow equation is

$$dC_t = dY_t + \tau Y \delta K_t dt + r(1 - \tau_C)C_t dt + dF_t - dD_t. \tag{4}$$

Holdings increase with the firm’s free cash flow, $dY$, and depreciation tax shield, $\tau Y \delta K$. The risk-free rate of return the firm earns on its uninvested resources, $r(1 - \tau_C)$, reflects a penalty on cash holdings. Funds can additionally be acquired through external financing or distributed to investors at any time.

Let $F_t$ and $D_t$ denote the cumulative (nondecreasing) funds acquired and dispensed by the firm up to time $t$ and $dF_t$ and $dD_t$ as the respective incremental changes in these policies over the time interval $(t, t + dt)$. When financing externally and receiving funds from financial markets, firms face explicit underwriting costs and implicit costs, as investors naturally question managers’ intended use of funds and the potential change to their incentives.

Modeling these costs are complicated but to provide an environment in which we can calibrate a model we follow Gomes (2001) who summarizes the costs of external financing with a fixed cost $\Phi$ and a marginal cost $\phi$. Together these costs imply that firms will only intermittently tap markets for funds and, when they do, raise a finite amount. To ensure firms do not outgrow financing costs, we follow Bolton et al. (2011) and assume both costs scale with capital as informational or incentive costs or the effects of dilution are likely to be proportional to firm size. We denote the cumulative costs of external financing up until time $t$ by $X_t$ and its incremental change as $dX_t$.

Because financing costs scale with capital, our model provides a better approximation to the behavior of large firms, as information asymmetries between insiders and outsiders are likely greater and vary more among small firms that often have shorter track records. Moreover, we calibrate issuance costs to equity and not debt markets, as equity issuances are more likely to be informationally sensitive. Homogeneity, importantly, makes the model tractible and Eberly, Rebelo and Vincent (2009) provide empirical support for it among large firms in Compustat.

Refinancing is not the only decision investors can make for they can also choose to let the firm fail and liquidate it. To distinguish these cases, we now turn to describing the contract between investors and managers.

---

6We do not solve for the optimal external financing policy jointly with the optimal incentive contract. Rather, incentives are made compatible given this particular institutional structure of financing costs.
C. INVESTORS’ AND MANAGEMENT’S CONTRACT AND TERMINATION

Investors hire a management team to run the firm and write a contract that can be terminated at any time. Investors have unlimited wealth and are risk neutral and therefore discount at the risk-free rate $r$. Managers are also risk neutral but discount at rate $\gamma > r$. They have no initial wealth and limited liability so investors cannot pay negative wages to them. At the termination time $\tau$ investors receive a fraction of assets and cash: $0 < l_K, l_C \leq 1$ of capital and resources, altogether recovering $l_K K_\tau + l_C C_\tau$. Managers receive their outside option, normalized to zero.

When management exerts no effort ($e_t = 0$) they enjoy private benefits at rate $\Lambda(K_t, C_t) dt$. Managers deriving benefits from capital, $\Lambda(K, \cdot)$, agrees with the long literature on empire building. There are several reasons for why they would also be expected to grow in resources, $\Lambda(\cdot, C)$. First, cash may provide funds for managers to invest in projects which offer private benefits but do not contribute to shareholder value. Because effort improves the return of productive assets ($E_e [dA_t] = e_t \mu dt$), no effort here can be viewed as putting effort into valueless projects that only managers enjoy. Second, more often than not, a cash-rich company runs the risk of being prodigal. And finally, large cash holdings remove some pressure on management to perform. Ultimately, that this function increases in both arguments is consistent with the thesis of Jensen (1986) and the international evidence on cash holdings and agency conflicts in Dittmar, Mahrt-Smith and Servaes (2003).

We assume the capital stock $K_t$, cash $C_t$, and cumulative free cash flow $Y_t$ are observable and contractible. From (1) and (2), investment $I_t$ and cumulative productivity $A_t$ can therefore be contracted upon. Investors maximize firm value by offering a contract that specifies investment, refinancing, and payout policies, $\{I\}$, $\{F\}$, and $\{D\}$, management’s cumulative payments, $\{U\}$, and a termination (stopping) time, $\tau$, all of which depend on the entire history of productivity $A_t$. Limited liability requires $U$ to be nondecreasing. We let $C = (I, F, D, U, \tau)$ represent the contract.

Given the contract, management chooses an effort process $\{e_t \in \{0, 1\} : 0 \leq t < \tau\}$ to solve

$$W(C) = \max_{\{e_t \in \{0, 1\} : 0 \leq t < \tau\}} E^e \left[ \int_0^\tau e^{-\gamma t} (dU_t + \Lambda(K_t, C_t)(1 - e_t) dt) \right],$$

where $E^e[\cdot]$ is the expectation operator under the probability measure induced by their effort choices. Their expected utility is composed of the present discounted value of compensation and

---

7This traditional assumption captures either their assumed impatience or in reduced-form the presence of outside investment opportunities available to them. While $\gamma = r$ may be a more neutral assumption, DeMarzo and Sannikov (2006) argue a contract can be made more robust by having investors assume that $\gamma$ is higher than managers’ true $\gamma$.

8Losing a fraction of resources is consistent losses in bankruptcy. In the broader Donaldsonian interpretation of resources, losses could reflect managers’ network or specific knowledge of the inner workings of the firm. Shue (2013) documents the importance of executive peers within a MBA cohort in explaining firm policies that do not necessarily contribute to firm productivity.
private benefits only when taking action $e_t = 0$.

At the time the contract is initiated, the firm has $K_0$ units of capital and $C_0$ units of cash. Given an initial payoff $W_0$ to managers, the problem investors face is

$$P(K_0, C_0, W_0) = \max_C \mathbb{E} \left[ \int_0^\tau e^{-rt}(dD_t - dF_t - dX_t - dU_t) + e^{-rt} (l_K K_t + l_C C_t) \right]$$

s.t. $C$ is incentive compatible and $W(C) = W_0$. (6)

The value to investors is the expected present discounted value of payouts, $dD$, less funds injected $dF$ at cost $dX$ and payments to managers $dU$, plus what they recover in liquidation.

Management’s payoff $W_0$ is determined by their relative bargaining power (DeMarzo and Sannikov (2006)). If managers possess all power, then $W_0^M \equiv \max \{ W : P(K_0, C_0, W) \geq 0 \}$. If however investors have all power, $W_0^I \equiv \arg\max_{W \geq 0} P(K_0, C_0, W)$. More generally, we blend the two extremes with a parameter $\psi \in (0, 1)$ by setting $W_0 = \psi W_0^M + (1 - \psi) W_0^I$.

### C.1. Incentive Compatible Contract

We focus on the case where the contract is incentive compatible and implements the efficient action $e_t = 1$ for all $t$. Given this contract and history up until time $t$, management’s continuation payoff is given by

$$W_t(C) = \mathbb{E}_t \left[ \int_t^\tau e^{-\gamma(s-t)}dU_s \right].$$

(7)

Note the collapse of the expectation operator $\mathbb{E}^*[\cdot]$ to the one that agrees with investors’ expectation $\mathbb{E}[\cdot]$. Rational expectations on behalf of both parties has management internalizing investors’ expectations and investors offering a contract consistent with those expectations. This model feature makes it dynamically consistent and formalizes the notion of market discipline affecting management’s behavior ex ante.

Standard dynamic contracting theory decomposes management’s incremental total compensation at time $t$ into incremental payments, $dU_t$, and incremental continuation payoff, $dW_t$ (Spear and Srivastava (1987) and Sannikov (2008)). The optimal contract compensates management for their time preference on average; the analogous promise keeping condition is $\mathbb{E}_t [dW_t + dU_t] = \gamma W_t dt$. Furthermore to maintain incentive compatibility, management’s compensation must remain sufficiently sensitive to the firm’s free cash flow, $dY_t$. Following DeMarzo and Sannikov (2006), we

---

Notice that payments to managers $dU$ are not subtracted from $dC$ in (4). We thoroughly discuss this alternative setup in Appendix [A]. In our setup, we are implicitly assuming that firms do not refinance to pay managers current payments, like bonuses.
formulate this sensitivity, $\beta_t$, with the martingale representation theorem:

$$dW_t + dU_t = \gamma W_t dt + \beta_t (dY_t - \mathbb{E}_t [dY_t]) = \gamma W_t dt + \beta_t (1 - \tau_Y) \sigma K_t dZ_t. \quad (8)$$

Agents who deviate reduce their compensation by $\beta_t (1 - \tau_Y) \mu K_t dt$ and receive private benefits $\Lambda (K_t, C_t) dt$. Incentive compatibility is thus implemented with $\beta_t (1 - \tau_Y) \mu K_t \geq \Lambda (K_t, C_t)$. Because liquidation is ex post inefficient and therefore costly to enforce, the optimal contract minimizes the likelihood of this event and sets

$$\beta_t = \frac{\Lambda (K_t, C_t)}{(1 - \tau_Y) \mu K_t} \text{ for all } t. \quad (9)$$

Intuitively, the optimal sensitivity is a ratio of private benefits to capital’s expected return potential and provides a nexus among free cash flow, compensation, capital, and cash holdings.

An important assumption is that the shocks to agents’ continuation utility are all local; that is, for a given $C_t$, $K_t$, and $W_t$, the instantaneous forward distribution of across two different economies are identical. Of course, firm value is generally not, and therefore dynamics will in general differ. This assumption is how we choose to model market discipline as an indirect force that invisibly guides managers’ behavior.

Given the functional form of $\Lambda (\cdot)$ it naturally implies that for a given beneficial shock to free cash flow ($dZ_t > 0$), greater capital or resources are associated with larger increases in managers’ compensation, consistent with Edmans, Gabaix and Landier (2008). In contrast, a series of negative shocks will result in either refinancing or else the termination of the contract and liquidation of the firm. For example if incentives have become too poor, investors will decide not to refinance the firm and let it fail. We now formalize this important decision while discussing the solution to our model.

II. Model Solution

Here we describe some properties of the solution to (6). Management’s continuation payoff $W_t$ in (7) is a state variable that summarizes management’s current incentives that reflect their expected path of compensation and the likelihood of contract termination. Capital $K_t$ captures the history of investment via (1). The firm’s cash holdings $C_t$ track the histories of refinancing and payouts. Altogether, whatever the history of the firm up until date $t$, the only relevant state variables are $K_t$, $C_t$, and $W_t$ and, therefore, investors’ value function at time $t$, $P(K_t, C_t, W_t)$, can be solved with a Hamilton-Jacobi-Bellman (HJB) equation.

We assume that adjustment costs, $G(I, K)$, and private benefits $\Lambda(K, C)$ are homogeneous of degree one in their arguments. The total cost of investment is thus $I + G(I, K) = Kg(i)$, where
\( i = I/K \) is the investment rate, and we specify private benefits to take a linear form

\[
\Lambda(K, C) = K\lambda(c) = K(\lambda_K + \lambda_C c)
\]  

(10)

that separates the agency friction attributed to capital (\( \lambda_K \)) and resources (\( \lambda_C \)). Homogeneity allows us to reduce the problem to two endogenous state variables—managers’ stake (their scaled continuation payoff), \( w = W/K \), and scaled cash holdings, \( c = C/K \)—and write \( p(c, w) = P(K, C, W)/K \).

Common to risk neutral models, investors optimally choose investment to equate expected returns to their required rate of return, the risk-free rate:

\[
rdt = \max_i \mathbb{E}_t \left[ \frac{d(Kp(c, w))}{Kp(c, w)} \right],
\]  

(11)

subject to the incentive compatibility constraint \( \beta \geq \lambda(c)/((1 - \tau Y)\mu) \) from (9) and (10). This equation’s solution is jointly determined with the boundaries that, when present, determine refinancing, payouts to investors, payments to managers, and contract termination.

In what follows we first present the solutions to models nested by our complete model as they are simpler yet still informative. We begin with the first-best model, then the model without an agency conflict followed by one without costly refinancing, before building to our complete model.

A. First-Best Solution

In the first-best economy there are neither agency (\( \Lambda(\cdot) = 0 \)) nor financial ((\( \Phi, \phi \)) = (0, 0)) frictions, management always chooses to exert effort, and the firm holds no cash and pays free cash flow out immediately. Because the economic environment is \( iid \) and the model homogeneous, there is a constant investment rate that maximizes firm value\(^{10}\)

\[
q^{FB} = \max_i \frac{(1 - \tau Y)\mu - g(i)}{r + \delta - i}.
\]  

(12)

In this economy, the classic Hayashi (1982) result equates average \( Q \) and marginal \( q \) to investment’s marginal cost to solve for optimal investment:

\[
g'(i^{FB}) = q^{FB} = \frac{(1 - \tau Y)\mu - g(i^{FB})}{r + \delta - i^{FB}}.
\]  

(13)

Finally, because our managers are relatively impatient, it is best to pay them immediately,

\(^{10}\)We assume \( \mu < g(r + \delta) \) and \( q^{FB} > I_K \) to have a well-defined problem.
leaving $P^{FB}(K, W) = q^{FB}K - wK$ to investors.

**B. Cash Management With No Agency Conflict**

In the absence of an agency friction ($\Lambda(\cdot) = 0$), management will always choose to exert effort. The only state variable is scaled cash holdings, $c = C/K$, implying firm value per unit of capital, $p(c)$, only depends on $c$, and the setup is similar to Bolton et al. (2011). We follow their exposition when it is always optimal to refinance the firm.

Because of the fixed issuance cost, the firm will want to minimize instances of refinancing and will therefore only do so when resources reach zero. When refinancing, the firm receives a total issue amount of $f > 0$ per unit of capital. Because firm value is continuous before and after issuance, value matching at the refinancing boundary $c = 0$ holds:

$$p(0) = p(f) - \Phi - (1 + \phi)f. \quad (14)$$

The right side is the firm’s post-financing value less fixed issuance costs $\Phi$ and proportional financing costs $\phi$. Because $f$ is optimally chosen, smooth pasting equates the marginal value of the last dollar raised $p'(f)$ to one plus the marginal financing cost,

$$p'(f) = 1 + \phi. \quad (15)$$

Conversely, because holding cash in the firm is penalized at rate $\tau_C$, the firm will distribute it to investors when abundant. Formally, let $\bar{c}$ denote this endogenous payout boundary and for $c > \bar{c}$ we have the equation $p(c) = p(\bar{c}) + (c - \bar{c})$. Because this equation holds continuously, the limit $c \to \bar{c}$ is summarized by the derivative

$$p'(\bar{c}) = 1. \quad (16)$$

Intuitively, at $\bar{c}$ the firm is indifferent between distributing and retaining one dollar, so the marginal value of cash must equal one. Since the payout boundary is optimally chosen, we also have the super contact condition holding at this point

$$p''(\bar{c}) = 0. \quad (17)$$

To summarize, incremental payouts $dD$ occur when $c \geq \bar{c}$ and incremental financing $dF$ is received when $c = 0$. Within these boundaries both $dD$ and $dF$ are zero and the dynamics of $\bar{c}$.
and (4) imply, by Ito’s lemma, that the evolution of scaled cash holdings is

\[ dc_t = [(1 - \tau_Y)\mu - g(i_t) + \tau_Y\delta + (r(1 - \tau_C) - (i_t - \delta))c_t]dt + \sigma(1 - \tau_Y)dZ_t. \]  

The solution to investors’ problem in (11) is in turn determined by the ordinary differential equation

\[ rp(c) = \max_i p(c)(i - \delta) + p'(c)[(1 - \tau_Y)\mu - g(i) + \tau_Y\delta + (r(1 - \tau_C) - (i - \delta))c] \\
+ \frac{1}{2}p''(c)\sigma^2(1 - \tau_Y)^2 \text{ for } c \in [0, \bar{c}] \]  

subject to the boundaries (14), (15), (16), and (17) that jointly pin down the location and optimality of the refinancing and payout decisions concerning \( f \) and \( \bar{c} \). The optimal rate of investment takes the form

\[ g'(i) = \frac{p(c)}{p'(c)} - c \]  

where its marginal cost is equated to its marginal benefit; namely, average \( Q \), \( p(c) \), adjusted for the marginal value of cash \( p'(c) \), less the reduction in cash holdings \( c \). If cash is scarce, its marginal value is high and then, for a given \( Q \), investment is diminished. Cash holdings thus influence the choice of investment.

C. AGENCY PROBLEM WITH COSTLESS REFINANCING

In the presence of an agency conflict yet the absence of external financing costs \( ((\Phi, \phi) = (0, 0)) \), episodes of refinancing are left indeterminate. Because cash held in the firm would incur a penalty, free cash flow is immediately paid to investors. The only state variable becomes managers’ stake, \( w = W/K \), making firm value per unit of capital to investors equal to \( p(w) \). The setup is then similar to DeMarzo et al. (2012) and following them the firm is liquidated when the contract is terminated.

Management will be terminated once their continuation utility hits zero (their outside option) because otherwise they would immediately consume private benefits. Hence, investors’ liquidation payoff at this termination boundary is

\[ p(0) = l_K. \]  

Next, because investors can always compensate management with cash, it will cost at most one dollar to increase \( w \) by one dollar, implying \( p_w(w) \geq -1 \). But because termination is costly
ex post it will be optimal to grow \( w \) at low values as quickly as possible by setting incremental payments \( dU/K \) in (8) to zero. As management is more impatient however \( (\gamma > r) \), at some point they will need to receive current payments. Formally, this payment boundary is the threshold where investors are indifferent between reducing their value by one dollar to pay agents one dollar immediately

\[
p'(w) = -1,
\]

and because it is determined optimally it satisfies the super contact condition

\[
p''(w) = 0.
\]

To summarize, \( dU_t/K_t = 0 \) within the payment and termination boundaries and \( \beta_t = \lambda_K/((1 - \tau_Y)\mu) \) for all \( t \). The evolution of \( w \) that is derived from the optimal contract then follows from (1) and (8):

\[
dw_t = (\gamma - (i - \delta))w_t dt + \frac{\sigma}{\mu} \lambda_K dZ_t.
\]

With the dynamics of management’s payoff given, the solution to investors’ problem in (11) is concave (see DeMarzo et al. (2012) for details) and takes the form

\[
 rp(w) = \max_i (1 - \tau_Y)\mu - g(i) + \tau_Y \delta + p(w)(i - \delta) + p'(w)(\gamma - (i - \delta))w
\]

\[
 + \frac{1}{2}p''(w) \left( \frac{\sigma}{\mu} \lambda_K \right)^2
\]

for \( w \in [0, w] \)

that is pinned down by the boundaries (21), (22), and (23) and where investment is determined by

\[
g'(i) = p(w) - p'(w)w.
\]

When choosing investment, investors internalize its effect on managers’ incentives. For a given \( w = W/K \), an increase in capital reduces management’s effective claim on the firm and induces a more severe agency friction.

D. Complete Model

We now discuss the complete solution in (11). It nests the previous two models as special cases and for brevity we streamline its presentation and draw attention only to novelties. Common to the nested models, within the boundaries of the solution the firm finances itself internally and only
affects management’s incentives through their continuation payoff. Total firm value per unit of capital now depends on both scaled cash holdings and managers’ stake, $p(c, w) + w$.

**D.1. Boundaries**

As before, the optimal contract specifies termination when managers’ stake equals their outside option

$$p(c, 0) = l_K + l_C \times c, \text{ for all } c. \tag{27}$$

Thus regardless of the level of cash the firm will be liquidated when $w = 0$ because at this point managers will shirk ($e_t = 0$) and investors will terminate the contract.

At stakes above $w = 0$, firm value satisfies $p_w(c, w) \geq -1$ and, again, at the payment boundary investors will be indifferent between promising and paying managers one dollar

$$p_w(c, \bar{w}(c)) = -1, \text{ for each } c, \tag{28}$$

while the super contact condition determines the level of the boundary itself

$$p_{ww}(c, \bar{w}(c)) = 0, \text{ for each } c. \tag{29}$$

We emphasize that the payment boundary $\bar{w}(c)$ is now a function: investors will choose to raise or lower the threshold depending on the level of cash. For example, if cash holdings increase from a small amount, the value of the firm will rise as the likelihood of costly refinancing falls. Because firm value is now higher, it will be optimal to reduce the probability of inefficient termination. Thus when starting at a low cash level, investors will find it efficient to further raise $\bar{w}(c)$ along with $c$ and shrink the likelihood of termination.

Next, we turn to the boundaries that determine the decisions of refinancing and payouts to investors. As before the firm refines only when it runs out of cash, but the magnitude, $f(w) > 0$, is now state-dependent.\footnote{We can show refinancing is always optimal at zero cash holdings. When holdings become zero, the liquidation value of the firm is $l_K - w$. Because $p_w(0, w) \geq -1$ over $w \in [0, \bar{w}(0)]$ and with equality when $w = \bar{w}(0)$, we have $p(0, w) = p(0, 0) + \int_0^w p_w(0, w')dw' \geq p(0, 0) + \int_0^w [1]dw' = p(0, 0) - w = l_K - w.$} To see this, since firm value is continuous before and after equity issuance it implies that

$$p(0, w) = p(0, 0) + \int_0^w p_w(0, w')dw' \geq p(0, 0) + \int_0^w [1]dw' = p(0, 0) - w = l_K - w. \tag{30}$$
where smooth pasting satisfies

\[ p_c(f(w), w) = 1 + \phi, \text{ for each } w. \]  \hspace{1cm} (31)

The size of refinancing now depends on management’s stake, \( w \). It is natural to believe that investors would likely provide more funds to a firm with a shrinking agency conflict. This intuition implies that we should expect to see regions where \( f'(w) > 0 \).

Because holding cash is costly, the firm will pay out once holdings are sufficiently large. Since firm value must be equal before and after a payout, the exact amount obeys the equation \( p(c, w) = p(\bar{c}(w), w) + (c - \bar{c}(w)) \) for \( c > \bar{c}(w) \), where we can see the payout boundary \( \bar{c}(w) \) is now a function of \( w \). And because it is continuous, the equation holds in the limiting case as \( c \to \bar{c}(w) \) and implies the boundary

\[ p_c(\bar{c}(w), w) = 1, \text{ for each } w, \]  \hspace{1cm} (32)

and again optimality requires

\[ p_{cc}(\bar{c}(w), w) = 0, \text{ for each } w. \]  \hspace{1cm} (33)

Altogether, the payout boundary becomes dependent on management’s stake, and the relation is not necessarily monotone for the following reasoning. It is possible in equilibrium that if \( w \) is sufficiently high it could be efficient to lower the cash threshold for payouts, conceivably making a costly refinancing event more likely. Of course, at high \( w \) investors know managers incentives are well aligned and will be motivated to keep refinancing a distant event. This effect arises because our economy is dynamically consistent and is also novel to the complete model.

And finally, we require along the boundary curves \( \bar{c}(w) \) and \( \bar{w}(c) \) that super contact holds with respect to both states. Economically, this means that along a segment of \( \bar{c}(w) \) it is optimal to not give managers current payments and similarly that along a part of \( \bar{w}(c) \) that investors will not earn payouts. Of course, there could be segments on which \( \bar{w}(c) \) and \( \bar{c}(w) \) overlap where both managers receive payments and investors payouts and form a joint upper boundary. These last few technical conditions follow below.

To ensure that \( \bar{c}(w) \) achieves super contact we first differentiate \( p(c, w) = p(\bar{c}(w), w) + (c - \bar{c}(w)) \) with respect to \( w \) which gives

\[ p_w(c, w) = \frac{\partial p(\bar{c}(w), w)}{\partial w} - \frac{\partial \bar{c}(w)}{\partial w}, \text{ for each } c \geq \bar{c}(w). \]  \hspace{1cm} (34)

Since the equation’s right side is not a function of cash, taking a derivative with respect to \( c \) and
letting \( c \to \bar{c}(w) \) implies

\[
p_{wc}(\bar{c}(w), w) = 0, \text{ for each } w. \tag{35}
\]

Next, a similar idea starting from \( p(c, w) = p(c, \bar{w}(c)) - (w - \bar{w}(c)) \) but differentiating with respect to \( c \) and then \( w \) and letting \( w \to \bar{w}(c) \) gives

\[
p_{cw}(c, \bar{w}(c)) = 0, \text{ for each } c. \tag{36}
\]

And finally from equations (35) and (36) it is evident that

\[
p_{cw}(\bar{c}(w), \bar{w}(c)) = 0, \text{ for every } c \text{ and } w. \tag{37}
\]

D.2. Summary

To summarize, events of refinancing \( dF \), payouts \( dD \), and payments to managers \( dU \) are zero within the boundaries and termination occurs when \( w = 0 \) regardless of cash holdings. Here, system dynamics are governed by

\[
dc_t = ((1 - \tau_Y)\mu - g(i_t) + \tau_Y \delta + [r(1 - \tau_C) - (i_t - \delta)]c_t)dt + \sigma(1 - \tau_Y)dZ_t \tag{38}
\]

and

\[
dw_t = (\gamma - (i_t - \delta))w_tdt + \sigma \frac{\mu}{\lambda(c_t)}dZ_t. \tag{39}
\]

Both resources and incentives vary with productivity and the optimal contract makes these two variables perfectly correlated. Because the drifts will differ in general, however, the stationary distribution will display the rich tradeoffs of the economic environment.

One such tradeoff arises from the interdependence of (38) and (39). The optimal contract sets \( \mu e_t = \mu \) implying that cash will grow quickly. But as cash holdings accumulate, so does the severity of the agency friction, \( \lambda(c) \). This raises the sensitivity of management’s compensation to the underlying productivity shocks. So while growth in productivity benefits cash holdings and potentially delays a refinancing event, the likelihood of contract termination increases for a given \( w \), making the firm riskier. Thus riskier firms will generally hold more cash, consistent with the empirical evidence in Acharya, Davydenko and Strebulaev (2012).

In what follows we assume that \( \beta p_{ww}/2 + p_{cw} \) and \( p_{ww} \) are nonpositive, conditions which we discuss further and verify numerically in Appendix A. The solution to (\Pi\Pi), then, can be repre-
presented by the partial differential equation

\[
rp(c, w) = \max_i p(c, w)(i - \delta) + p_c(c, w)((1 - \tau_Y)\mu - g(i) + \delta\tau_Y + [r(1 - \tau_C) - (i - \delta)]c) \\
+ p_w(c, w)((\gamma - (i - \delta))w) + \frac{1}{2}p_{cc}(c, w)(\sigma(1 - \tau_Y))^2 \\
+ \frac{1}{2}p_{ww}(c, w)\left(\frac{\sigma}{\mu}\lambda(c)\right)^2 + p_{cw}(c, w)\frac{\sigma^2(1 - \tau_Y)}{\mu}\lambda(c). \tag{40}
\]

subject to incentive boundaries in (27), (28), and (29) that determine termination and the payment threshold to management, \(\bar{w}(c)\); the resource boundaries of (30), (31), (32), and (33) that locate the position and ensure the optimality of the refinancing and payout decisions, \(f(w)\) and \(\bar{c}(w)\); as well as the mixed boundaries given by (35), (36), and (37). We detail our computational method to solve this problem in Appendix A. A key novelty relative to Achdou, Han, Lasry, Lions and Moll (forthcoming) is that the shape of state space is solved jointly with (40).

The optimal investment decision is determined by

\[
g'(i) = \frac{p(c, w) - p_w(c, w)w}{p_c(c, w)} - c \tag{41}
\]

and now reflects several margins: the gain in value, \(p(c, w)\), less the detrimental change to managers’ incentives, \(p_w(c, w)w\), adjusted for the marginal value of cash, \(p_c(c, w)\), and its reduction, \(c\). More generally, the decision deepens the link between cash holdings, compensation, and investment. Empirically, increasing long-term incentive plans (LTIP) raises investment (Larcker (1983) and Glover and Levine (2017)) as do cash holdings, a result more or less established in Fazzari, Hubbard and Petersen (1988) and subsequently refined by a large literature.

**D.3. Distortions and The Joint Upper Boundary**

A key tradeoff balances greater cash holdings \(c\) with a greater agency friction \(\lambda(c)\) while considering both parties’ interests. Ideally, the firm would locate where both financing and agency frictions and their value distortions are minimized, which respectively correspond to where \(p_c(\bar{c}(w), w) = 1\) and \(p_w(c, \bar{w}(c)) = -1\).

A remarkable outcome is that in spite of the problem’s complexity, our setup produces a simple, intuitive tradeoff along the joint optimal boundary, a fact which we summarize in the following proposition:

**Proposition (Tradeoff Along the Joint Upper Boundary).** Consider a marginal change along the joint upper boundary from \((\bar{c}(w), \bar{w}(c))\) to \((\bar{c}(w) + dc, \bar{w}(c) + dw)\), then the rate of change across
this boundary is equal to

\[ \frac{dw}{dc} = -\frac{r\tau_C}{\gamma - r} < 0 \]  

(42)

Proof. See Appendix A.

At the boundary at which investors receive payouts and managers payments, the slope equals the ratio of marginal costs of retaining cash to withholding payments to managers. Even though the value distortions have dissipated \((p_c(\cdot) = 1 \text{ and } p_w(\cdot) = -1)\), the lingering sources of inefficiency, the cash tax penalty and managers’ impatience, survive and determine the optimal tradeoff faced by firms.

More specifically, if investors decide to hold an additional amount \(dc\) of cash, they will bring managers’ payment boundary inwards by \(\frac{r\tau_C}{\gamma - r} dc\). Accordingly, the proposition suggests that the rates and magnitudes at which managers or investors are differentially paid are informative about the relative size of these underlying frictions.

More generally, movements away from the joint upper boundary either increase cash’s marginal value, \(p_c(\cdot) > 1\) as costly refinancing becomes more likely, or raise the marginal value of an additional dollar promised to managers, \(p_w(\cdot) > -1\), as it becomes more valuable to avoid costly termination. Therefore as either \(c\) or \(w\) decrease, the marginal costs of not paying out shareholders or not currently paying managers fall and change the tradeoff faced along the boundary.

Mathematically this means the curvature of the boundary on either side of the joint upper boundary reflects marginal changes in relative distortions. Shifts away from \((\overline{c}(w), \overline{w}(c))\) and along \((\overline{c}(w), w)\), for example, are informative about the distortion attributed to the agency friction. The model can novelly be used to measure these distortions.

In our specified setup, the tradeoff is exactly linear at the joint upper boundary and a different specification, like decreasing returns to scale, could relax this linearity. The general insight that the curvature of the boundary is informative about the underlying magnitudes of distortions, however, would remain.

D.4. Aggregation

With the description of firm behavior complete, we now describe the stationary distribution of firms. While the payout, payment, and refinancing boundaries are not absorbing, the termination boundary is. Because of this, every firm will eventually fail and in order to study a stationary distribution we therefore allow entry. The exit rate, moreover, is an salient equilibrium object that governs the severity of the agency friction.

Each firm is described by its current state \((c, w)\), and therefore the density of firms is defined over this state space. The non-stationary distribution at time \(t\), \(h(c, w, t)\), satisfies the Kolmogorov
forward equation

\[
\frac{\partial h(c, w, t)}{\partial t} = \varphi(c, w)m + A^*h(c, w, t),
\]

(43)

where \( A^*h(c, w, t) \) is the adjoint of the infinitesimal generator of the bivariate diffusion process \((dc_t, dw_t)\).\(^{12}\) By construction, this generator contains the rates of exit that occur along the termination boundary \( w = 0 \). To ensure a stationary mass of firms, we add a product of an entry rate \( m \) and an entry mass \( \varphi(c, w) \) that integrates to one.

We pin down the entry rate in the stationary distribution with the normalization that the total mass of firms is a constant equal to one:

\[
\int_{c}^{\tau(w)} \int_{w}^{\pi(c)} h(c, w) dw dc = 1.
\]

After this normalization notice the left side of (43), twice integrated, is zero and we can then rearrange it for the stationary entry rate, which by construction equals the exiting mass of firms,

\[
m = - \int_{0}^{\tau(w)} \int_{0}^{\pi(c)} A^*h(c, w) dw dc.
\]

(44)

When a firm’s contract is terminated, a new, replacing firm’s cash holdings is drawn from a distribution with positive support. The entrant, however, also starts with a new continuation payoff, \( w_0 \), that is determined by the bargaining power of agents and investors. We specify initial conditions during our calibration in the next section.

III. Model Calibration and Analysis

Having characterized the solution to the model, we now calibrate it before turning to study the solution’s properties. After the calibration and as before, we build to the complete model by revisiting the predictions of its nested models. We then define the measurement of financial and agency distortions.

A. Calibration

Our calibration is summarized in Table I. It is split into externally- and internally-calibrated parameters that target informative data moments. Our empirical environment contains only US public firms, as agency frictions are likely to be present among them, and report the details of our widely-used Compustat and Execucomp data samples in Appendix B.

We begin by calibrating our external parameters by setting the tax rate on corporate income to \( \tau = 30 \) percent, the interest rate to \( r = 4 \) percent, and the depreciation rate of assets to \( \delta = 8 \).
percent, all common values in the literature. We choose to model a tax penalty for holding cash in the firm. Two relevant sections of the IRS tax code are Section 531 on the accumulated earnings tax and Section 541 on undistributed personal holding company income. Both sections impose the same penalty rate and we accordingly use $\tau_C = 20$ percent.

We specify a smooth adjustment cost technology as we are interested in the model’s long-run properties that takes the quadratic form

$$g(i) = i + \frac{\theta}{2}(i - \delta - z)^2,$$

(45)

where $\theta$ measures the magnitude of the adjustment cost and $z$ is an exogenous expansion rate that locates the function. Following Hall (2001), we interpret the parameter $\theta$ as a doubling time of capital. He uses either 2 or 8 years for his upward adjustment cost and 20 or 80 years for his downward adjustment cost. Because financial and agency frictions will lower investment rates below first-best and potentially the depreciation rate of capital, we assume an exogenous expansion rate $z$ to match the average investment rate in the data and generate long-run capital growth. Our choices are $\theta = 6$ and $z = 9$ percent.

Next, we turn to the costs of refinance. Beginning with the seminal work of Jensen and Meckling (1976) and Myers and Majluf (1984), subsequent literature has tried to estimate indirect costs, like asymmetric information and incentive costs, and direct costs, underwriting fees and dilution for example. Estimates vary across studies: Calomiris and Tsoutsoura (2013) argue a 3 percent decline in the price of equity in response to a seasoned equity offering is reasonable but can be as high as 15 percent for smaller firms in totality when accounting for all costs; and Altinkiliç and Hansen (2000) show that the majority of costs for a seasoned offering are variable, ranging from 4 to 6 percent depending on issue size, with fixed costs slightly below half a percent. Informed by them, we impose $\Phi = 0.5$ percent and $\phi = 5$ percent.

In the event of contract termination, we assume the firm is liquidated. In a recent study of recovery rates within bankruptcies, Kermani and Ma (2020) estimate values between 33 and 46 percent for all assets. Chen (2010) structurally estimates average recovery rates for bondholders to near 40 percent. We assume a $l_K = 40$ percent recovery rate for capital and the perfect recovery of cash, $l_C = 1$.

A.1. Internal Calibration: Averages

We internally calibrate our remaining six parameters $(\mu, \sigma, \gamma, \lambda_K, \lambda_C, \psi)$ to features of the stationary distribution that clearly map model to data. This distribution encodes all information about optimal policy functions and is therefore an ideal target for calibration. Specifically, we target moments of free cash flow, compensation, and cash holdings as well as the shape of its state space.
that determine the frequencies and magnitudes of payouts, refinancing, and termination.

The incremental return on capital, $\mu$, directly influences the mean rate of free cash flows. We set $\mu = 0.17$ to match this average.

Specialists’ time rate of preference is $\gamma > r$. Its value influences the length of the interval $[0, \overline{w}(c)]$, as greater impatience (higher $\gamma$) requires sooner current payments and lowers $\overline{w}(c)$. In reality managerial compensation, while easily measurable, is complex as it contains salary, variable bonuses, long-term incentive plan contributions, and stock and options, the timing of which can also follow a complicated structure. As a resolution, we convert the stock of expected discounted future compensation, $W$, to a flow by multiplying it by $\gamma$, avoiding the subjective calculation necessary to evaluate (7) and effectively using the flow of compensation in the data to match the optimal contract’s promising-keeping condition and track management’s stake in the firm. Altogether, we set the parameter by calibrating to average compensation in the data and correspondingly choose $\gamma = 0.048$.

Finally, we set the agency costs across both capital and resources as $\lambda_K = 0.04$, which influences the variation in management’s continuation utility and the probability of contract termination, and $\lambda_C = 0.09$, which influences cash’s marginal value and therefore average cash holdings.

Recall that failure is source of inefficiency in the model and provides discipline to managers who under individual rationality would prefer to remain in the firm. The average termination rate in the model is $m$ from (44). We target this rate with $\sigma = 35$ percent to match the two percent average default rates of public firms over 1993 to 2017 (Boualam, Gomes and Ward (2020)). Thus our volatility parameter is targeted at the frequency of events which are determined by the boundaries rather than, say, the cross-sectional dispersion in investment rates.

We factor the entry mass into a conditional and marginal distribution, $\varphi(c_0, w_0) = \varphi_c(w_0|c_0)\varphi_c(c_0)$ and assume initial scaled cash holdings, $c_0$, draws from a log-normal distribution with mean $0.15 - \sigma^2/2$ and standard deviation $\sigma$, which generates an entrant’s cash holdings close to the model’s average refinancing size. Given the cash draw, $c_0$, the distribution of $w_0$ is degenerate and the value of initial $w_0$ comes from an assumption of managers’ relative bargaining power. From wage responses to news, Taylor (2013) structurally estimates relative bargaining power to be equally split between shareholders and the chief executive and, accordingly, we pick $\psi = 50$ percent.

### A.2. Internal Calibration: Frequencies and Magnitudes

Next, we describe how we construct the model’s frequencies and magnitudes of refinancing and payouts to closely match their empirical construction. In the data, we form an indicator for a firm for whether it had ever, over the course of an entire year, had refinanced or paid out and simply average over these indicators to estimate frequencies. We do not record multiple events of the same
firm within a year.

Our motivation for calculating the payout frequency in the model, then, follows from the question: Given the stationary mass at a point, \((c, w)\), what fraction of firms would be expected to breach the payout boundary following a productivity shock, \(\Delta z\), given at an annual rate? Given the shock, cash holdings and managers’ stake move by \(\Delta c = \mu_c(c, w) + \sigma_c \Delta z\) and \(\Delta w = \mu_w(c, w) + \sigma_w(c) \Delta z\), respectively, where \(\mu_c(c, w), \sigma_c, \mu_w(c, w),\) and \(\sigma_w(c)\) are the annualized drifts and volatilities of (38) and (39). Given these moves, we calculate

\[
\text{Refinancing Rate: } E[1 \{ \Delta c + c < 0 \} | c, w] = \mathcal{N}\left(-\frac{\mu_c(c, w) + c}{\sigma_c}\right),
\]

\[
\text{Payout Rate: } E[1 \{ \bar{c}(w + \Delta w) < \Delta c + c \} | c, w] = E \left[ 1 \{ \Delta z > (\bar{c}(w + \mu_w(c, w) + \sigma_w(c) \Delta z) - c - \mu_c(c, w))/\sigma_c\} \right],
\]

where \(\mathcal{N}(\cdot)\) is the cumulative standard normal distribution and we compute the rate of payouts with Gaussian quadrature while accounting for the correlation between \(c\) and \(w\). Refinancing size is simply \(E[f(w)|w]\) and is consistent with our empirical construction. These objects are conditional on \((c, w)\) and we integrate over them with the stationary density to calculate refinancing and payout statistics.

### A.3. Internal Calibration: Summary

The summary of the internal calibration is tabulated in Table II. The model matches well the data’s average levels of cash, compensation, investment, free cash flow, and entry/exit and refinancing rates.

One feature of the data that the model has difficulty in matching is refinancing size. When it occurs in the data, it raises a much larger amount on average than in the model. Small firms are known to raise a lot more upon refinancing (Fama and French (2005)) and so allowing for decreasing returns to scale would help the model in this dimension.

The payout rate of the model, moreover, is lower than in the data. Payouts in the data are measured using common dividends and repurchases. In the model, payouts are more akin to special one-time dividends and repurchases, and so its frequency will naturally be lower since dividend policies are known to be quite persistent (Lintner (1956)). Additionally, the decision to return cash to investors would depend on the rate of return to the firm’s investment. Here, decreasing returns to scale might also help as they would cause larger firms to have relatively lower investment returns and make payouts more appealing.
B. Solution to Cash Management with No Agency Conflict

Figure 1 plots the solution to the optimal cash management policy without an agency conflict being present. In Panel A, enterprise value is plotted in blue, \( p(c) - c \), over the domain of scaled cash holdings, \( c = C/K \in [0, \bar{c}] \), where \( \bar{c} \) is the payout boundary. The red dashed line depicts marginal financing costs, \( \Phi + \phi c \), and its tangency with enterprise value determines the refinancing size \( f \).

Because refinancing is always preferred to liquidation, firm value is above \( l_K \) when cash is zero. It is concave in this region, reflecting the precautionary motive induced by the prospect of costly refinancing. As \( c \) grows and eventually reaches \( \bar{c} \), the slope of enterprise value becomes zero. Concurrently, cash’s marginal value is initially above one but converges to one.

Panel B plots the investment rate, \( i = I/K \), as a function of cash holdings. The decision is influenced by the marginal value of cash that reflects financial frictions. As this marginal value grows the firm reduces investment to tilt the drift of cash’s evolution upwards and reduce the likelihood of costly refinancing.

The stationary density of cash holdings is depicted in gray. Intuitively, the mass of firms above the refinancing size \( f \) dominates the mass below and cluster near the upper boundary. Refinancing is costly and firms avoid this by accumulating cash and making investment decisions to promote this goal.

C. Solution to Agency Problem with Costless Refinancing

Figure 2 depicts the solution to the agency model in the absence of external financing costs. Panel A shows the value function of investors, \( p(w) \), under a contract of commitment and over the domain of managers’ stake, \( w = W/K \in [0, \bar{w}] \), where \( \bar{w} \) is the payment boundary.

Because termination is ex post inefficient, investors become averse to fluctuations in \( w \). Investor’s value function, \( p(w) \), is thus concave as investors internalize the risk that the optimal contract places on management’s actions. At \( w = 0 \) the contract is terminated and investors receive the liquidation value \( l_K \). Upon termination, a new firm is drawn depending on managers’ bargaining power, \( \psi \).

Two forces drive the shape of \( p(w) \). Initially as \( w \) grows from zero, the agency friction falls and agrees with investors’ desire to weaken it to avoid liquidation, \( p(w) \) thus increases. To further alleviate the agency friction, investors must promise a larger and larger share of the firm to management, and therefore \( p(w) \) eventually declines reflecting this wealth transfer. As \( w \) increases, the slope of the value function declines and eventually becomes \(-1\) at the payment boundary, the point at which investors are indifferent between promising and paying managers one dollar.

Panel B plots the investment rate. To understand this graph it is best to think of the firm’s total value being determined in part by investors’ capital, \( K \), and in part by management’s ability to run
the firm and their stake in the firm, $W$. Investment is costly to investors as it lowers management’s effective ownership of the firm, $w = W/K$, and hence induces a greater agency friction. The cost of a greater friction is much larger at low levels of $w$ and this is why investment is reduced. As the friction wanes with growing $w$, investment waxes along with it.

Finally, the stationary density of management’s stake is shown in gray. Because liquidation is costly, the optimal contract sets $du = 0$ to grow $w$ as quickly as possible. Because of management’s impatience, however, $du$ is eventually set to be positive at the payment boundary, $\overline{w}$. The density accumulates at this upper bound.

D. SOLUTION TO COMPLETE MODEL AND ITS PROPERTIES

The complete model retains the properties of the two nested models previously described. Investment, for example, is generally reduced as $w$ or $c$ fall from the payment or payout boundaries. The optimality of its solution similarly requires investors’ value function to possess certain properties prescribed by the equilibrium contract as well as the decisions that determine the boundary conditions. In particular, the derivatives of the value function to each state, $w$ and $c$, should be monotone decreasing functions, respectively, and the second derivatives should equal zero at the upper boundaries.

We depict the accuracy of these properties in Figure 3. Panels A and B report the first own-derivatives of a state across three percentiles of the other state’s marginal density. Both the marginal cost of compensation $p_w(c, w)$ and the marginal value of cash $p_c(c, w)$ are decreasing functions, implying own-state concavity of the value function. As they approach their respective boundaries, the rate of change of these derivatives fall and approach zero. A notable difference between the complete model and the agency model with costless refinancing is that $p_w(c, w) < 0$ for all $w$ here and therefore the contract is renegotiation-proof.

Next, in Panels C and D we plot the super contact condition associated with the payout and payment boundaries. If the decision is optimal, they should both be uniformly zero across the entire boundary. In general they are very close, although in the tails of the distribution of the state in question the magnitude of the deviation from zero grows. These deviations visually overstate the impact on the model’s predictions, as they are concentrated over states on which the equilibrium stationary distribution puts little mass, as depicted by the marginal densities in gray. Sensitivity analysis confirms that the quantitative predictions of the model are robust to local changes in boundary curves.

Figure 4 displays the complete solution. Panel A shows the value function that is solved jointly with the state space. The boundary touches both zero axes for cash holdings and managers’ stake. Remaining on a curve while moving in to the interior, the boundary curve eventually reaches the joint upper boundary, marked by a black square. The rate of change of either boundary, $\tau(w)$ or
\( \overline{w}(c) \), reflects the current marginal values of states, \( p_c(c, w) \) and \( p_w(c, w) \), that in turn reflect the underlying financial and agency frictions.

Intuitively, the rise of \( c \) from zero coincides with a higher firm value and a greater likelihood of termination, making it efficient to raise the payment boundary \( \overline{w}(c) \). At some point, however, the cost of termination falls as cash holdings and liquidation value have grown, bringing the payment boundary back in. In addition, as \( w \) moves away from zero, firm value increases and reduces the probability of termination. The expected return on investing capital within the firm is thus higher and improves the motive to push the payout boundary out farther. When managers’ incentives have improved enough, both party’s interests are well-aligned, and it becomes optimal to lower the payout boundary, reflecting our dynamically consistent contract whereby managers share investors interests. In general, the overall shape is determined by these competing effects.

The refinancing curve, \( f(w) \), is plotted as a dotted black line. This function captures the disciplinary role of markets on management. In equilibrium, management understands that refinancing will be dependent on the firm’s history up to that point and therefore this will change their behavior ex ante. As expected, refinancing does increase with \( w \). As \( w \) rises from zero, \( f(w) \) initially grows rapidly, but then plateaus and eventually falls.

Why it falls is reminiscent of Zwiebel (1996), who argues that managers voluntarily set debt to restrict themselves. Analogously, in our model management internalizes refinancing’s effect on their incentives. They realize that with already good incentives, a larger issue causes unnecessarily greater risk to investors as \( \lambda(c) \) and the risk of termination grow with \( c \) for a given \( w \), and refinancing \( f(w) \) thus falls. In general, the quantitative effects of increasing \( f(w) \) and managers voluntarily restraining themselves determine the appearance of the curve.

The stationary density of firms is graphed in Panel B. The density approximately tracks out a path from zero cash holdings to the payout boundary, a track far above the termination boundary. Its particular gradient arises from the equilibrium investment decisions that drive the state equations. Because the model is stochastic the mass of firms are spread out along this track, causing some firms to fail.

E. MEASUREMENT OF DISTORTIONS

The traditional approach to measuring distortions typically compares market values and investment across first-best and second-best outcomes. Market values, however, rely on an accurate discounting of future cash flows that are very influenced by hard-to-measure discount rates. Fully understanding investment, moreover, requires a good proxy for marginal \( q \), an object notoriously hard to estimate (Erickson and Whited (2000)). First-best, furthermore, may not be a reasonable benchmark as it is unlikely to be attained in practice as financial and agency frictions do exist.

The Proposition of Section II defines the slope of the joint upper boundary and implies the
existence of a second-best frontier that is tangent to this boundary. We therefore use the proposition to measure distortions relative to this frontier and form measurements based on quantities rather than on prices.

We construct an orthogonalized definition of the agency distortion by evaluating the distance between the frontier and the payout boundary, weighted by the mass of firms across this distance. We use a similar definition for the financial distortion. We denote the frontier as a function of state by $F(c)$ and $F^{-1}(w)$ and in Appendix A we derive the formal definitions

\[
\text{Agency Distortion: } \mathbb{E}[\Delta c 1\{\bar{c}(w) < c + \Delta c < F^{-1}(w)\}|c, w],
\]

\[
\text{Financial Distortion: } \mathbb{E}[\Delta w 1\{\bar{w}(c) < w + \Delta w < F(c)\}|c, w].
\]

Similar to the frequency calculations from before, they are based on the annualized drifts and volatilities of (38) and (39) and, as they are functions of $(c, w)$, we integrate over them with the stationary density to calculate the economy’s average agency and financial distortions.

Why do we measure the agency distortion by the distance from the payout boundary? Compare a firm at the joint upper boundary $(\bar{c}(w), \bar{w}(c))$ to one on $(\bar{c}(w), w)$ where $w < \bar{w}(c)$. For this firm, the financial friction is due only to the holding cash penalty, $r \tau C$, since $p_c(\bar{c}(w), w) = 1$, so the only distortion from the second-best frontier can be attributed to agency. The idea is akin to holding using a supply shock to identify a demand elasticity.

Put differently, payouts which occur between $\bar{c}(w)$ and $F^{-1}(w)$ are required by investors to compensate them for holding a firm with an acute agency friction. Analogously, current payments which occur between $\bar{w}(c)$ and $F(c)$ are promised to managers to operate a firm with a scarce cash holdings. These average distortions thus provide an economic interpretation of how much investors are compensated on average to remain invested in an agency-laden firm or managers to remain operating a cash-poor one.

Figure 5 illustrates the approach to measuring the agency distortion. It depicts the top-down view of the stationary density. The second-best frontier, the dashed-dot line, bounds the state space from above and produces a tangency at the joint upper boundary marked by the black square. Given a firm at $(c, w)$ and the dynamics of (38) alone (recall our distortions’ definitions are orthogonalized), we can evaluate the probability that this firm will receive a shock and cross the payout boundary $\bar{c}(w)$. For a given $w$, we truncate the likelihood of shock realizations beyond $F^{-1}(w)$ as they surpass second-best outcomes. Since the shock changes cash holdings (or compensation), we are measuring quantities.

To sum up, we provide novel, theoretically-consistent measurements of financial and agency distortions and show that payouts and compensation, two readily observable variables, can evaluate their magnitudes. Common to the literature in general, the measurements of these distortions are
conditional on a model, as they are in Hsieh and Klenow (2009), and an alternative model would lead to different estimates. In Hsieh and Klenow (2009), they partially control for this shortcoming by comparing the same model across countries and evaluate relative distortions. A different yet also valid approach would be to pick a model and compute the ratio of payouts to payments. This ratio mimics an index of the relative severity of agency to financial distortions and, usefully, can be constructed for one industry in one country. Evaluating changes in this index are more robust and less affected by model misspecification.

IV. MODEL ANALYSIS AND EVALUATION OF MARKET DISCIPLINE

With all objects defined and the model solved, we now focus on two forms of model analysis. First, we contrast impulse response functions across different calibrations. Recall that market discipline is an indirect force which we model by having only local shocks affect management’s continuation utility in (8). Dynamic analysis allows us to visualize and quantify this force by showing that firms, even with the same initial \((c, w)\), evolve differently over time.

Second, we use steady state analysis pioneered in Hopenhayn (1992) by observing how the stationary density of firms shifts in response to a change in a model parameter. We use it here to understand how changes in industrial structure, whether in the tax code or in the severity of a deep agency friction, affect the observable characteristics of firms and the distortions present in the economy. Though these stationary densities remain constant through time, they do so by the neutralizing effects of entry and exit, of firm growth and contraction. This analysis is therefore useful for understanding adaptation in ex ante behavior as it captures the long-run effects of these structural changes.

A. EVALUATION OF MARKET DISCIPLINE

We assess market discipline’s quantitative effect in the following way. We first solve two economies that differ by a parameter. For illustration we contrast the fixed cost of refinance, setting \(\Phi\) to 50 and 150 basis points. We then choose an initial pair of states that defines a firm, \((c_0, w_0)\) and let investment, scaled cash holdings, and managers’ stake evolve in the absence of future shocks and track the evolution of a firm’s policy. Related, an impulse response function maps the dynamics of firm policies from the steady state after experiencing a normalized shock, typically. Because our steady state is composed of a distribution of firms, our notion here is the natural extension of an impulse response function to an arbitrary firm in the steady state distribution.

Our firm is initialized with scaled cash holdings near the average refinancing amount in data, \(c_0 = 15\) percent of net assets, and a level of managerial compensation equal to the average in the data, \(\gamma w_0\), equal to 1.3 percent of net assets. The response functions are depicted in Figure 6.
As expected, both scaled cash holdings and managers’ stake start at the same level. Over time, however, they diverge because firm value and the state space differ.

Cash holdings are more quickly accumulated in the high $\Phi$ economy, ending up nearly 5 percentage points higher. Managers, however, end up with a smaller ownership stake in the firm. As cash is more valuable in the high $\Phi$ economy, the probability of refinancing within the year falls, but the likelihood of payouts is largely unaffected. Effectively managers, not investors, bear the change in fixed cost. That in turn predicts that distortions attributed to agency are now greater and basically only become quantitatively similar after 20 quarters, when nearing the joint upper boundary. Finally, financial distortions are roughly similar, as the effects of a greater fixed cost of refinance have been offset by a more rapid accumulation of cash.

Altogether, a small change in fixed cost of refinance can have a dramatic impact on managers’ compensation, firms’ cash holdings, and the agency distortion in the economy. This underscores that market discipline has a material force on economic dynamics.

B. Steady State Analysis

The results of this form of analysis are summarized in Table III. For convenience, the first column restates the benchmark moments targeted in the internal calibration. It also includes estimates of average distortions and the ratio of their magnitude relative to the benchmark.

Our benchmark calibration has firms paying out 2.54 percent of firm assets per year out as a result of agency frictions. Managers are compensated by an additional 0.27 cents per dollar of assets to operate riskier firms with low cash ratios. We find agency frictions to be nearly ten times more severe than financial ones when measured with our model’s quantities. The intuition is that financial frictions can simply be offset by accumulating cash (Bates, Kahle and Stulz (2009) document an abnormally high accumulation rate since the 1990s). An accumulation of cash, however, is double-edged as it exacerbates the agency conflict, which requires payouts to investors (see Farr-Mensa, Michaely and Schmalz (2014) for corroborating empirical evidence connecting payouts and agency).

The scenario in Column (2) raises average productivity, $\mu$, to 18 percent. The firm now as a whole is more profitable and managers are correspondingly paid more and terminated less. The return to investment rises and the firm holds more cash, reducing the frequency of external finance. These patterns are consistent with a boom. In this economy, cash holdings and capital grow, implying that so do agency frictions. The model therefore predicts agency conflicts to be procyclical and financial frictions countercyclical.

In Column (3) we lower the corporate tax rate from 30 to 21 percent, matching a change enacted in 2017. The tax cut raises average asset productivity and strengthens the precautionary savings motive for holding cash. Although reducing corporate taxes is widely believed to lead
to stimulating investment, we find that the effect is only modest as it raises the investment rate not even one percentage point. Investors instead prefer to allocate the additional free cash flow to managerial compensation to alleviate agency frictions and to distributing cash to themselves.

The change of the fixed refinancing cost, $\Phi$, appears in column (4). Frequencies of payout and refinancing fall, similar to the boom-like patterns of column (2), but markets refinance on more stringent terms and as a result they terminate more firms. Greater cash holdings effectively offset a potentially aggravated financial friction. But managers must now operate more cautiously and with more resources that can potentially be squandered. Investors thus require more compensation for a larger agency conflict, as can be seen by an agency distortion growing from 2.54 to 2.92.

**B.1. A Preliminary Proposal to Reduce Agency Frictions**

Columns (3) and (4) collectively mimic a corporate income tax cut combined with the introduction of a tax of one percent on the instance of refinancing. We combine both changes in column (5). As before, a higher $\Phi$ reduces refinancing rates and increases the probability of termination, a change which alone would require a greater compensation for agency conflicts demanded by investors. Combining this change with lower corporate taxes, however, raises average cash flows that allows investors to reallocate a portion of them towards rewarding managers and alleviating agency conflicts.

In the final column (6) we initially lower the agency friction attributed to cash from 0.09 to 0.05 and then lower managers’ bargaining power to match the entry/exit rate of column (6). Several rows across columns (5) and (6) look similar quantitatively. And relative to the benchmark case the frequency of refinancing has fallen and conditional on it happening, its average size has grown.

Altogether, an economy that implements the tax proposal above generates an economy that mimics one with a relatively less severe agency friction. This is an imputed result of market discipline. The relative values of average distortions are near identical.

In its current state, however, the analysis shown here is only suggestive and presents a tradeoff whereby the overall effects on agency and financial frictions must be weighed. Of course a fuller and potentially general equilibrium analysis that includes a government budget constraint would be required to be more confident in prescriptions for policy. But we find it interesting nonetheless and leave a more analysis to future work that we discuss next in the conclusion.

**CONCLUSION**

We quantitatively evaluate the fundamentally important question of the degree to which investor and managerial incentives are aligned and the role markets play in attaining firm value maximization. We formalize the notion of market discipline whereby markets, even though tapped
intermittently, invisibly guide management’s use of resources.

Our quantitative model clarifies the role of markets in affecting a wide range of firm policies, from cash holdings, investment, payouts, compensation, to whether to refinance a firm or let it fail. We also derive a novel, general formula that shows how investor payouts and managers’ compensation are informative about the underlying distortions of costly external finance and agency conflicts. Our benchmark calibration has firms paying out 2.54 percent of firm assets per year out as a result of agency frictions and managers are compensated by an additional 27 cents per dollar of assets to operate riskier firms with low cash ratios, implying agency frictions are nearly 10 times more severe than financial ones.

While novel, our analysis necessarily omits some features that we believe are important. First is decreasing returns to capital. The model-data fit on statistics of refinancing and payouts would be expected to improve with this amendment. Yet another useful extension would be clearly formulate capital structure. Our current setup works best to describe large firms that are not overly indebted. Other improvements that could be equally important include time-variation in aggregate states. Lustig, Syverson and Van Nieuwerburgh (2011) partially attribute the rise in the disparity across executive compensation to changes in executive’s outside options. Bolton, Chen and Wang (2013) entertain a model where market conditions fluctuate and influence the costs of external financing over time. One last extension would be to model takeovers, board composition, and competition, as these likely influence managers’ behavior. We leave these variations on our benchmark model to future work.


Farre-Mensa, Joan, Roni Michaely, and Martin Schmalz, *Payout Policy*, Vol. 6,


A. TECHNICAL APPENDIX

A. DETAILS OF SOLUTION METHOD

We solve the partial differential equation in (40) with a finite difference method that approximates the function \( p(c, w) \) on a two-dimensional non-rectangular grid: \( c \in \{ c_i(w_j) \}_{i=1}^{I} \) and \( w \in \{ w_j(c_i) \}_{j=1}^{J} \), where we define \( \bar{w}(c_i) = w_{j}(c_i) \) and \( \bar{c}(w_j) = c_{i}(w_j) \). Each set of grid points along \( j \), \( w_j(c_i) \), depend on the value of \( c_i \), because of the boundary curve \( \{ \bar{w}(c_i) \}_{i=1}^{I} \). The set of grid points along \( i \), \( c_i(w_j) \) shares the same logic.

We approximate first derivatives of \( p \) using both backward and forward differences and second derivatives with central differences. All differences of \( c \) and \( w \) are calculated respectively over the fixed increments \( \Delta_c \) and \( \Delta_w \). For the approximation of the derivatives at the boundaries, there are three different cases:

1. The boundary conditions of \( w \) imply that \( p(c, 0) = l_K + l_C c \Rightarrow p(c_i, w_0) \approx l_K + l_C c_i \) and \( p_w(c, \bar{w}(c)) = -1 \Rightarrow p(c_i, w_{j+1}) \approx p(c_i, w_{j}) - \Delta_w \) under a forward difference, where both conditions hold for all \( i \).

2. The boundary conditions of \( c \) imply that \( p(0, w) = p(f(w), w) - \Phi - (1 + \phi)f(w) \Rightarrow p(c_0, w_j) \approx p(f(w_j), w_j) - \Phi - (1 + \phi)f(w_j) \) and \( p_c(\bar{c}(w), w) = 1 \Rightarrow p(c_{i+1}, w_j) \approx p(c_i, w_{j}) + \Delta_c \) under a forward difference, where both conditions hold for all \( j \).

3. The boundary conditions along the joint upper boundary where \( p_{cw}(\bar{w}(w), \bar{w}(c)) = 0 \) for all \( \bar{w}(w) \) and \( \bar{w}(c) \) implies

\[
p(c_{i+1}, w_{j+1}) \approx p(c_{i+1}, w_{j}) - \Delta_w \approx p(c_i, w_{j}) - \Delta_w + \Delta_c.
\]

We describe our computational algorithm below:

1. Guess the value of \( p^b (c, w) \) on the two-dimensional non-rectangular grid: \( c \in \{ c_i \}_{i=1}^{I} \) and \( w \in \{ w_j(c_i) \}_{j=1}^{J} \) and approximate the derivatives

2. Calculate the investment policy function in (41)

3. For each \( w \) in \([w_1(c_1), w_{J1}(c_1)]\), we use bisection to find the refinancing policy \( f(w) \) such that \( p_c(f(w), w) = 1 + \phi \)

4. We update the value function through an implicit method that solves the vector

\[
p^{b+1} = (p_{1,1}^{b+1}, \ldots, p_{1,J1}^{b+1}, p_{2,1}^{b+1}, \ldots, p_{2,J2}^{b+1}, \ldots, p_{I,J}^{b+1})' \]

with notation \( p_{i,j} = p(c_i, w_j) \). It begins with a guess \( b = 1 \) and proceeds to iterate until convergence (\( \max(|p^{b+1} - p^b|) < 10^{-9} \)) on the value function

\[
p^{b+1} \left[ \left( \frac{1}{\Delta} + r - (i - \delta) \right) - Q \right] = p^b / \Delta + B,
\]

where \( i \) is calculated from step 3, \( \Delta > 0 \) is the step size of the iterative method, and \( Q \) is the transition
matrix defined by the diffusion processes of the states $c$ and $w$ and the boundaries described above.

$$ Q = \begin{bmatrix} q_{1,1}^{ss} & q_{1,1}^{su} & 0 & \cdots & 0 & q_{1,1}^{us} & q_{1,1}^{uu} & 0 & \cdots & 0 & \cdots & 0 \\ q_{1,2}^{sd} & q_{1,2}^{ss} & q_{1,2}^{su} & \cdots & q_{1,2}^{us} & q_{1,2}^{uu} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & q_{1,j_1}^{sd} & q_{1,j_1}^{ss} & 0 & \cdots & q_{1,j_1}^{us} & q_{1,j_1}^{uu} & \cdots & \cdots & \cdots & \cdots \\ q_{2,1}^{ds} & q_{2,1}^{du} & 0 & \cdots & 0 & q_{2,1}^{su} & q_{2,1}^{uu} & 0 & \cdots & 0 & \cdots & \cdots \\ q_{2,2}^{dd} & q_{2,2}^{ds} & q_{2,2}^{du} & \cdots & q_{2,2}^{us} & q_{2,2}^{uu} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & q_{2,j_2}^{dd} & q_{2,j_2}^{ds} & 0 & \cdots & q_{2,j_2}^{us} & q_{2,j_2}^{uu} & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & q_{j_1,j_1}^{sd} & q_{j_1,j_1}^{ss} \\ \end{bmatrix} $$

Adjustments to transition rates along the boundaries are made to $Q$ for the non-rectangular grid as it is an approximation and ensure that the non-termination-boundary rows of the transition matrix $Q$ sum to zero. The termination-boundary rows do not sum to zero as they measure the (absorbing) exiting mass of firms. The matrix $Q$ is the discretized analogy of the infinitesimal generator of $(dc_t, dw_t)$: $A\vartheta(c, w)$ for some arbitrary function $\vartheta(\cdot)$. The elements of $Q$ are based on an upwind scheme and defined as

- $q_{i,j}^{ss} = \max(\mathbb{E}_t[dc], 0) / \Delta_c + \min(\mathbb{E}_t[dc], 0) / \Delta_c - \mathbb{E}_t[dc] \cdot \Delta_c / \Delta_c$,
- $q_{i,j}^{su} = \min(\mathbb{E}_t[dw], 0) / \Delta_w + \mathbb{E}_t[dc] \cdot \Delta_c / \Delta_c$,
- $q_{i,j}^{sd} = -\min(\mathbb{E}_t[dc], 0) / \Delta_c + \mathbb{E}_t[dc] \cdot \Delta_c / \Delta_c$,
- $q_{i,j}^{us} = \max(\mathbb{E}_t[dc], 0) / \Delta_c + \mathbb{E}_t[dc] \cdot \Delta_c / \Delta_c$,
- $q_{i,j}^{dd} = -\min(\mathbb{E}_t[dc], 0) / \Delta_c + \mathbb{E}_t[dc] \cdot \Delta_c / \Delta_c$,
- $q_{i,j}^{du} = -q_{i,j}^{ud} = q_{i,j}^{dd} = \mathbb{E}_t[dc] \cdot \Delta_c / (4 \Delta_c \Delta_w)$,

where the conditional moments of state variables are $\mathbb{E}_t[dc] = (\gamma - (i - \delta))w$, $\mathbb{E}_t[dc] = ((1 - \tau) - \mu + g(i) + \delta \tau Y + [r(1 - \tau C_1 - (i - \delta))]c$), $\mathbb{E}_t[dc^2] = (\sigma \lambda(c) / \mu)^2$, $\mathbb{E}_t[dc^2] = (\sigma (1 - \tau))^2$, and $\mathbb{E}_t[dc dw] = \sigma^2 (1 - \tau Y) \lambda(c) / \mu$. 

36
Lastly, the vector of constants $B$ required by the boundaries takes the form

$$B = \begin{bmatrix}
(q_{11}^{dd} + q_{11}^{id} + q_{11}^{ud}) \times (l_K + l_C c_1)
\vdots
(q_{21}^{dd} + q_{21}^{id} + q_{21}^{ud}) \times (l_K + l_C c_2)
(q_{11,j1}^{ud}) \times (-\Delta_w)
\vdots
(q_{21,j2}^{ud}) \times (-\Delta_w)
\vdots
(q_{j1,j1}^{ud}) \times (-\Delta_w)
\end{bmatrix}$$

and intuitively captures the rates of cash outflows from payments to managers, $-\Delta_w$, and inflows to investors from payouts, $\Delta_c$, liquidation, $l_K + l_C c$, and refinancing, $p_{1,j}$ for $j = 1, \ldots, J^1$, where we use the equilibrium value matching condition for refinancing, \([30]\).

### A.1. Stationary Distribution

The stationary distribution, $h(c, w)$, is calculated by solving, $h(c, w) = -(Q^T)^{-1}\psi$, where $\psi$ is the entry vector. The rows of $\psi$ that are non-zero are determined by the assumed shape of the entry distribution that isolates $c$ and the assumption on how agents’ initial continuation utility $w$ is determined through bargaining power. The normalization of $h(c, w)$ to one implies that the entry rate equals $m = -\sum_j Q^T h(c, w) \Delta_w \Delta_c$.

In the model with no agency friction there is no exit as firms always refinance. Here we compute the stationary distribution $h(c)$ by solving an eigenvalue problem of the adjoint of the transition matrix $Q$: $Q^T h(c) = 0$.

### B. Proof of Tradeoff Along the Joint Upper Boundary

We first prove a lemma on investment being constant along the joint upper boundary before turning to prove the proposition. In what follows we ignore dependence of boundaries on states for brevity: that is, $\overline{c} = \overline{c}(w)$ and $\overline{w} = \overline{w}(c)$.

**Lemma.** Investment is constant along the joint upper boundary.

**Proof.** Denote $(\overline{c}, \overline{w})$ and $(\overline{c} + dc, \overline{w} + dw)$ as two points along the joint upper boundary. From the first-order condition in \([41]\) we have

$$g'(i(c, w)) = \frac{p(c, w) - p_w(c, w)w}{p_c(c, w)} - c,$$

where optimality implies that the investment rates at these two points are

$$g'(i(\overline{c}, \overline{w})) = p(\overline{c}, \overline{w}) + \overline{w} - \overline{c}, \quad \text{and} \quad A1$$

$$g'(i(\overline{c} + dc, \overline{w} + dw)) = p(\overline{c} + dc, \overline{w} + dw) + (\overline{w} + dw) - (\overline{c} + dc). \quad \text{(A2)}$$

Next, from the continuity of the value function along the boundary we know

$$\frac{p(\overline{c}, \overline{w} + dw) - p(\overline{c}, \overline{w})}{dw} = -1 \quad \text{and} \quad \frac{p(\overline{c} + dc, \overline{w} + dw) - p(\overline{c}, \overline{w} + dw)}{dc} = 1,$$

37
which we can rearrange to yield

\[ p(\bar{c}, \bar{w}) = p(\bar{c}, \bar{w} + dw) + dw = p(\bar{c} + dc, \bar{w} + dw) - dc + dw \]  
(A3)

and then adding \( \bar{w} - \bar{c} \) to both sides of (A3) gives

\[ p(\bar{c}, \bar{w}) + \bar{w} - \bar{c} = p(\bar{c} + dc, \bar{w} + dw) + (\bar{w} + dw) - (\bar{c} + dc). \]  
(A4)

Finally, the optimal investment rates along the boundary ((A1) and (A2)) and (A4) imply

\[ i(\bar{c}, \bar{w}) = i(\bar{c} + dc, \bar{w} + dw). \]

Therefore, the investment rate is constant along the joint upper boundary.

Now we turn to proving the proposition, which we restate here for convenience.

**Proposition (Tradeoff Along the Joint Upper Boundary).** *Consider a marginal change along the joint upper boundary from \((\bar{c}(w), \bar{w}(c))\) to \((\bar{c}(w) + dc, \bar{w}(c) + dw)\), then the rate of change across this boundary is equal to

\[ \frac{dw}{dc} = \frac{-r\tau_c}{\gamma - r} < 0 \]

*Proof.* The partial differential equation in (40) is

\[
 rp(c, w) = \max_i p(c, w)(i - \delta) + p_c(c, w) \left((1 - \tau_Y)\mu - g(i) + \delta\tau_Y + [r(1 - \tau_C) - (i - \delta)]c\right) \\
 + p_w(c, w)(\gamma - (i - \delta))w + \frac{1}{2} p_{cc}(c, w)(\sigma(1 - \tau_Y))^2 \\
 + \frac{1}{2} p_{cw}(c, w) \left(\frac{\sigma}{\mu} \lambda(c)\right)^2 + p_{cw}(c, w) \frac{\sigma^2(1 - \tau_Y)}{\mu} \lambda(c). 
\]

(A5)

Using our two points, \((\bar{c}, \bar{w})\) and \((\bar{c} + dc, \bar{w} + dw)\), we can reduce (A5) to

\[
 p(\bar{c}, \bar{w})[r - (i - \delta)] = -(\gamma - (i - \delta))\bar{w} \\
 + (1 - \tau_Y)\mu - g(i) + \delta\tau_Y + [r(1 - \tau_C) - (i - \delta)]\bar{c} 
\]

(A6)

and

\[
 p(\bar{c} + dc, \bar{w} + dw)[r - (i - \delta)] = -(\gamma - (i - \delta))(\bar{w} + dw) \\
 + (1 - \tau_Y)\mu - g(i) + \delta\tau_Y + [r(1 - \tau_C) - (i - \delta)](\bar{c} + dc),
\]

(A7)

where we use the Lemma to simplify investment \( i \equiv i(\bar{c}, \bar{w}) = i(\bar{c} + dc, \bar{w} + dw) \). Therefore, when subtracting (A6) from (A7) we are left with

\[
 [r - (i - \delta)]\left(p(\bar{c} + dc, \bar{w} + dw) - p(\bar{c}, \bar{w})\right) = [r(1 - \tau_C) - (i - \delta)]dc - (\gamma - (i - \delta))dw. 
\]

(A8)

We can then place (A3) into (A8) to find

\[
 [r - (i - \delta)](dc - dw) = [r(1 - \tau_C) - (i - \delta)]dc - (\gamma - (i - \delta))dw,
\]

which reduces to (42) in the text.  

\[ \square \]
C. DERIVATIONS OF DISTORTIONS

We derive the conditional expectation for the agency distortion and use but do not report for brevity a similar derivation for the financial distortion (49). We ignore the dependence of \( \mu_c(c, w) \) on state variables in what follows.

\[
\mathbb{E}[\Delta c \mathbb{1}\{\overline{c} < c + \Delta c < F^{-1}(w)\}|c, w] = \int_{-\infty}^{\infty} \Delta c \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{1}{2}(\frac{\Delta c - \mu_c}{\sigma_c})^2} \mathbb{1}\{\overline{c}(w) - c < \Delta c < F^{-1}(w) - c\} \, d(\Delta c) \tag{A9}
\]

We then use the change of variable \( \Delta x = (\Delta c - \mu_c)/\sigma_c \) to change (A9) to

\[
\int_{-\infty}^{\infty} \frac{\sigma_c}{\sqrt{2\pi}} e^{-\frac{1}{2}(\Delta x)^2} \mathbb{1}\{\overline{c}(w) - c < \sigma_c\Delta x + \mu_c < F^{-1}(w) - c\} \, d(\sigma_c\Delta x + \mu_c) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\Delta x)^2} \mathbb{1}\{\overline{c}(w) - c - \mu_c < \Delta x < \frac{F^{-1}(w) - c - \mu_c}{\sigma_c}\} \, d(\Delta x) = \int_{\frac{\overline{c}(w) - c - \mu_c}{\sigma_c}}^{\frac{F^{-1}(w) - c - \mu_c}{\sigma_c}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\Delta x)^2} \, d(\Delta x).
\]

Since under the change \( u = -\frac{1}{2}x^2 \) we have \( \int xe^{-\frac{1}{2}x^2} \, dx = -\int e^u du = -e^u + k = -e^{-\frac{1}{2}x^2} + k \), where \( k \) is a constant of integration, equation (A10) equals

\[
-\frac{\sigma_c}{\sqrt{2\pi}} e^{-\frac{1}{2}(\Delta x)^2} \mathbb{1}\{\overline{c}(w) - c - \mu_c\} + \mu_c \int_{\frac{\overline{c}(w) - c - \mu_c}{\sigma_c}}^{\frac{F^{-1}(w) - c - \mu_c}{\sigma_c}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\Delta x)^2} \, d(\Delta x) = \frac{\sigma_c}{\sqrt{2\pi}} \left[ -e^{-\frac{1}{2}((\overline{c}(w) - c - \mu_c)/\sigma_c)^2} - e^{-\frac{1}{2}(\overline{c}(w) - c - \mu_c)/\sigma_c} \right] + \mu_c \mathcal{N}\left(\frac{\overline{c}(w) - c - \mu_c}{\sigma_c}\right) - \mathcal{N}\left(\frac{F^{-1}(w) - c - \mu_c}{\sigma_c}\right).
\]

D. VERIFICATION OF HJB OPTIMALITY AND FULL EFFORT CONDITION

Define the gain process \( \{G\} \) under any incentive-compatible contract \( \mathcal{C} = (I, F, D, U, \tau) \) for any \( t \leq \tau \) as

\[
\mathcal{G}_t(\mathcal{C}) = \int_0^t e^{-rs}(dD_s - dF_s - dX_s - dU_s) + e^{-rt}P(K_t, C_t, W_t),
\]

where \( K_t, C_t, \) and \( W_t \) evolve as in (1), (4), and (8). Homogeneity and Ito’s lemma imply

\[
e^{rt} d\mathcal{G}_t = K_t \begin{bmatrix} -rp + p(i_t - \delta) + pc((1 - \gamma_t)\mu - g(i_t)) + \overline{\delta}(1 - \gamma_t) - (i_t - \delta)c_t) \\
+pwc(\gamma - (i_t - \delta)w_t + \frac{1}{2}pwc(\sigma(1 - \gamma_t))^2 + \frac{1}{2}pwc(\beta_t(1 - \gamma_t)^2 + pwc\beta_t(1 - \gamma_t)^2\sigma^2) + [1 - p_c] \times (dD_t - dF_t) - dX_t - (1 + p_w)dU_t) \end{bmatrix}
\]

where \( p(\cdot) \)’s dependence on states \( (c_t, w_t) \) and \( \mathcal{G}(\cdot) \)’s dependence on a contract \( \mathcal{C} \) have been omitted for conciseness from this point on.

Under the optimal investment \( i^*_t \), and incentive policies \( \beta^*_t = \lambda(c_t)/((1 - \gamma_t)\mu) \) the top two lines in
the square brackets are the optimized PDE in (40) and therefore zero. For models in which the only state variable is agents’ continuation utility this nonpositivity condition follows from the concavity of \( p(w) \). In this more general case, we show and verify numerically that for any other incentive-compatible policy both \( p_{uw} \) and the sum \( \beta p_{uw}/2 + p_{cw} \), under the policy with the smallest \( \beta \), are nonpositive. \(^{13}\) Panels A and B of Figure A-1 depict \( p_{uw} \) and the sum under \( \beta^*_\ast \) and the calibration in Table I.

The figure broadly shows that these terms are negative across the entire state space. However, in the upper and lower 10 percent of the distributions, indicated by the dashed lines, there are instances of these terms being slightly positive, especially in the lower portion of \( w \)’s distribution. Solving this model is difficult, particularly since the exact shape of the boundary of the state space is not well understood and functional forms must be used to approximate it. Accordingly, these instances coincide with regions where super contact conditions diverge further from zero (see Panels C and D of Figure 3) and are therefore likely due in part to numerical error and also coincide with regions where the stationary density does not place a large mass (see Panel B of Figure 4). Because of this latter fact, perturbing the boundary to minimize further this error has little impact on the quantitative predictions of our model. In our defense, there is no other evidence of non-optimality in the solution and we place no restrictions on agents’ ability to process information (for example, Krusell and Smith (1998)).

Next, the term capturing the optimality of the continuation payment policy, \(- (1 + p_w) du_t \), is nonpositive since \( p_w \geq -1 \) but equals zero under the optimal contract. The term that captures the optimality of net funds dispensed, \( (1 - p_c)(dD_t - dF_t) \), is also nonpositive since (i) \( p_c \geq 1 \) and (ii) \( dD_t \geq dF_t \) because either (ii.a) cash holdings are sufficient for payouts, \( dD > 0 \) and \( dF = 0 \); or (ii.b) if current cash holdings are insufficient to finance payouts the firm will raise funds externally for payouts, in which case \( dD \geq dF \) since cash holdings are positive. This term is also zero under the optimal contract. Lastly, the issuance cost, \(- dX_t \), is nonpositive but equals zero under the optimal contract.

Therefore, for the auxiliary gain process we have

\[
dG_t = \mu_G(t) dt + e^{-rt} K_t (p_c + \beta_t p_w) \sigma (1 - \tau_t) dZ_t,
\]

where \( \mu_G(t) \leq 0 \). Let \( \varphi_t \equiv e^{-rt} K_t (p_c + \beta_t p_w) \sigma (1 - \tau_t) \). We impose the usual regularity conditions to ensure that \( \mathbb{E} \left[ \int_0^T \varphi_t dZ_t \right] = 0 \) for all \( T > 0 \). This implies that \( \{G\} \) is a supermartingale.

We can now evaluate the principal’s payoff for an arbitrary incentive compatible contract. Recall that \( P(K, C, W) = l_K K + l_C C \). Given any \( t < \infty \),

\[
\begin{align*}
&\mathbb{E} \left[ \int_0^\tau e^{-rs} (dD_s - dF_s - dX_s - dU_s) + e^{-rt} \left( l_K K_s + l_C C_s \right) \right] \\
= &\mathbb{E} \left[ \mathcal{G}_{t\wedge \tau} + \int_{\tau}^\tau e^{-rs} (dD_s - dF_s - dX_s - dU_s) + e^{-rt} \left( l_K K_s + l_C C_s \right) - e^{-rt} P(K_t, C_t, W_t) \right] \\
= &\mathbb{E} \left[ \mathcal{G}_{t\wedge \tau} + e^{-rt} \left( \int_{\tau}^\tau e^{-r(s-\tau)} (dD_{s-\tau} - dF_{s-\tau} - dX_{s-\tau} - dU_{s-\tau}) + e^{-r(\tau-\tau)} \left( l_K K_{s-\tau} + l_C C_{s-\tau} \right) - P(K_{s-\tau}, C_{s-\tau}, W_{s-\tau}) \right) \right] \\
\leq &\mathcal{G}_0 + (q^{FB} - (l_K + l_C \times c)) \mathbb{E}[e^{-rt} K].
\end{align*}
\]

The first term of the inequality follows from the nonpositive drift of \( d\mathcal{G}_t \) and the martingale property of

\(^{13}\)This can be seen by simplifying the second-order terms that depend on the incentive coefficient \( \beta, p_{uw} \beta^2(1 - \tau_y) \sigma^2 / 2 + p_{cw} \beta(1 - \tau_y) \sigma^2 \), to be less than or equal to zero.
The second term follows from
\[
\mathbb{E}_t \left[ \int_t^T e^{-r(s-t)} (dD_s - dF_s - dX_s - dU_s) + e^{-r(t-s)} (l_K K_t + l_C C_t) \right] \leq q^{FB} K_t - w_t K_t,
\]
which is the first-best result and
\[
q^{FB} K_t - w_t K_t - P(K_t, C_t, W_t) \leq (q^{FB} - p(c, 0)) K_t = \left( q^{FB} - (l_K + l_C \times c) \right) K_t.
\]
as \( w + p(c, w) \) is increasing in \( w \) since \( p_w(c, w) \geq -1 \).

We impose the standard transversality conditions \( \lim_{T \to \infty} \mathbb{E}[e^{-rT} K_T] = 0 \) and \( \lim_{T \to \infty} \mathbb{E}[e^{-rT} C_T] = 0 \). Therefore letting \( t \to \infty \)
\[
\mathbb{E} \left[ \int_0^t e^{-rs} (dD_s - dF_s - dX_s - dU_s) + e^{-r(t-s)} (l_K K_t + l_C C_t) \right] \leq \mathcal{G}_0. \tag{A11}
\]
for all incentive-compatible contracts. On the other hand, under the optimal contract \( C^* \), principal’s payoff \( \mathcal{G}(C^*) \) achieves \( \mathcal{G}_0 \) because the above weak inequality holds with equality when \( t \to \infty \).

### D.1. Full Effort Condition

Finally, we require \( \Lambda(K_t, C_t) \) to be sufficiently small to ensure the optimality of \( e_t = 1 \) all the time. When managers shirk \( (e_t = 0) \) they enjoy private benefits at rate \( \Lambda(K_t, C_t) dt \). Cash holdings would then evolve as
\[
dC_t = (1 - \tau_Y) \sigma K_t dZ_t - I_t dt - G(I_t, K_t) dt + \tau_Y \delta K_t dt + r (1 - \tau_C) C_t dt + dF_t - dD_t.
\]
When they shirk their payoff would not depend on cash flow realizations, so their continuation payoff would change according to
\[
dW_t = \gamma W_t dt - dU_t - \Lambda(K_t, C_t) dt.
\]

For this not to be the case and for effort \( (e_t = 1) \) to remain optimal, it must be that investors’ payoff rate from allowing agents to shirk be lower than under the optimal contract and equivalently that investors’ optimal gain process remain a supermartingale with respect to this shirking policy:
\[
\begin{align*}
 rp & \geq p(i - \delta) + p_c (-g(i) + \delta \tau_Y + [r(1 - \tau_C) - (i - \delta)]c) \\
 & \quad + p_w (\gamma - \lambda(c) - (i - \delta)) w + \frac{1}{2} p_{cc} \sigma (1 - \tau_Y)^2, \text{ for all } c \text{ and } w.
\end{align*}
\]

We confirm this optimality of full effort numerically and depict the result in Panels C and D of Figure A-1. In more general situations where the inequality binds, then a more complicated contract than the one described in this paper would need to be considered. Zhu (2013) considers this extended contracting environment in the context of the DeMarzo and Sannikov (2006) model.

### E. Alternative Setup Where Managers are Paid Out of Cash

In this appendix we derive the necessary equations required for the optimality of the equilibrium under the assumption that managers’ incremental payments \( dU \) subtract from incremental cash holdings \( dC \). We streamline its presentation and elaborate after on the key differences from our benchmark setup.
Under this alternative setup, cash holdings possess the law of motion

\[ dC_t = dY_t + \tau_Y \delta K_t dt + r(1 - \tau_C)C_t dt + dF_t - dD_t - dU_t \]

and investors maximize

\[ P(K_0, C_0, W_0) = \max_c \mathbb{E} \left[ \int_0^\tau e^{-rt} (dD_t - dF_t - dX_t) + e^{-r\tau} (l_K K_\tau + l_C C_\tau) \right] \]

subject to similar conditions as in (6). The laws of motion for \( K \) and \( W \) are identical to the benchmark’s. The scaled HJB equation then becomes under \( dU = dF = dD = dX = 0 \) within the boundaries

\[
rp(c, w) = \max_i p(c, w)(i - \delta) + p_c(c, w)((1 - \tau_Y)\mu - g(i) + [r(1 - \tau_C) - (i - \delta)]c) + p_w(c, w)(\gamma - i + \delta)w
\]

\[
+ \frac{1}{2}p_{cc}(\sigma(1 - \tau_Y))^2 + \frac{1}{2}p_{ww}\left(\frac{\sigma}{\mu}\lambda(c)\right)^2 + p_{cw}(c, w)\frac{\sigma^2(1 - \tau_Y)}{\mu}\lambda(c),
\]

which is identical to the benchmark’s. We enumerate the boundary conditions below.

1. The termination boundary is

\[ p(c, 0) = l_K + l_C \times c \text{ for all } c. \]

2. For each \( c \), there is a compensation level \( \overline{w}(c) \) at which it is optimal to pay managers in current payments,

\[
\frac{p(c, w) - p(c, \overline{w}(c))}{w - \overline{w}(c)} = \frac{p(c - (w - \overline{w}(c)), \overline{w}(c)) - (w - \overline{w}(c))}{w - \overline{w}(c)}
\]

\[
\Rightarrow \frac{p(c, w) - p(c, \overline{w}(c))}{w - \overline{w}(c)} = \frac{p(c - (w - \overline{w}(c)), \overline{w}(c))}{w - \overline{w}(c)} - 1
\]

\[
\Rightarrow p_w(c, \overline{w}(c)) = -p_c(c, \overline{w}(c)) - 1, \text{ as } w \to \overline{w}(c), \quad (A12)
\]

and which requires the condition \( p_{ww}(c, \overline{w}(c)) = 0 \) for each \( c \).

3. When cash holdings reach zero, the firm refinances with an equity issue of size \( f K \). The refinancing decision is determined in part by where \( w \) lies relative to \( \overline{w}(f) \), managers’ payment boundary for each level of post-refinancing cash holdings.

- If \( w \in [0, \overline{w}(f)] \), then for each \( w \)

\[ p(0, w) = p(f, w) - \Phi - (1 + \phi)f, \quad (A13) \]

along with the condition \( p_c(f, w) = 1 + \phi \) as before.

- If \( w > \overline{w}(f) \), then the firm will refinance and pay management current payments, so for each \( w \)

\[ p(0, w) = p(f, \overline{w}(f)) - \Phi - (1 + \phi)f - (w - \overline{w}(f)). \quad (A14) \]

Differentiating this equation with respect to \( c \) gives

\[ p_c(f, \overline{w}(f)) + p_w(f, \overline{w}(f)) \frac{\partial \overline{w}(c)}{\partial c} \bigg|_{c=f} + \frac{\partial \overline{w}(c)}{\partial c} \bigg|_{c=f} = 1 + \phi. \quad (A15) \]
Because this first-order condition depends only on the marginal value of cash and does not depend on \( w \), we can use continuity of \( w \) approaching \( \overline{w}(c) \) in (A12) to write

\[
p_w(f, \overline{w}(f)) = p_w(c, \overline{w}(c))|_{c=f} = -1 - p_c(c, \overline{w}(c))|_{c=f} = -1 - p_c(f, \overline{w}(f))
\]

and combining with (A15) we get

\[
p_c(f, \overline{w}(f)) \left( 1 - \frac{\partial \overline{w}(c)}{\partial c} \bigg|_{c=f} \right) = 1 + \phi. \tag{A16}
\]

4. When cash holdings get large we have as before

\[p_c(\overline{w}(w), w) = 1 \text{ and } p_{cc}(\overline{w}(w), w) = 0 \text{ for each } w.\]

5. And finally the mixed derivatives along the boundaries require that

\[
p_{cw}(\overline{w}(w), w) = 0 \text{ for each } w,
\]

\[
p_{cw}(c, \overline{w}(c)) = -p_{cc}(c, \overline{w}(c)) \text{ for each } c, \text{ and}
\]

\[
p_{cw}(\overline{w}(w), \overline{w}(c)) = -p_{cc}(\overline{w}(w), \overline{w}(c)) = 0 \text{ for every } c \text{ and } w.
\]

We now discuss the salient differences between the alternative setup and our benchmark. From bullet 5., the mixed derivative at the payment boundary \( p_{cw}(c, \overline{w}(c)) = -p_{cc}(c, \overline{w}(c)) \) no longer equals zero because it now needs to account for the reduction in cash holdings.

From 2., \( p_w(c, w) + p_c(c, w) \geq -1 \) rather than simply \( p_w(c, w) \geq -1 \), showing the intuitive change that the bound of the marginal cost of compensation now depends on the marginal cost of cash. Along the payout boundary \( \overline{w}(w) \) the inequality collapses to \( p_w(\overline{w}(w), w) \geq -2 \): that is, it now costs investors at most two dollars to raise managers’ continuation utility marginally—the reduction in cash holdings costs investors one dollar \( (p_c(\overline{w}(w), w) = 1) \) and raising management’s continuation utility costs, at most, another dollar. Because (A12) implies that \( p_c(c, \overline{w}(c)) \) decreases in \( c \), the slope along the payment boundary, \( \partial \overline{w}(c)/\partial c \), should be negative.

At last from 3., refinancing can now follow two decision rules depending on the location of \( w \) relative to \( \overline{w}(f) \). This hypothetical decision is depicted in Figure A-2. The refinancing decision \( f(w) \) is traced out with the dashed line. Refinancing decisions satisfy (A13) as in our benchmark and is depicted by the arrow from \( p(0, w) \) to \( p(f, w) \), where \( f \) is determined by \( p_c(f, w) = 1 + \phi \). At some \( w \), however, it may be optimal to refinance and concurrently pay managers. This decision is depicted by the top two arrows, first moving from \( p(0, w) \) to \( p(f, w) \), reflecting the post-refinancing gain in value, and then downwards from \( w \) to \( \overline{w}(f) \), reflecting the transfer from investors to managers of size \( w - \overline{w}(f) \). Along this two-part arrow, the refinancing decision is determined by (A15). The firm would decide by choosing the maximum of (A13) and (A14).

Of course, this is not the only possibility for refinancing. A different equilibrium could be imposed by simply requiring (A13) to hold at all points in the alternative setup. Under this policy, refinancing and current payments to managers would never co-occur, as it is in our benchmark. However, in this alternative setup this decision may not be optimal on behalf of investors as they might prefer to refinance more and pay managers, whereas in our benchmark it is optimal.

To sum up, there are several reasons that we prefer our benchmark to this alternative setup. First, the HJB equations are identical. Second, the alternative setup introduces multiple choices in the refinancing decision and it is unclear a priori how to select the correct choices. Third, the refinancing region satisfying (A14) is unlikely to matter quantitatively, since the stationary distribution is likely to put effectively zero
mass on low cash, high manager stake firms because of the optimal contract placing a perfect correlation structure across $dc_t$ and $dw_t$. Altogether it introduces more complexity into an already challenging setup and is unlikely to make a quantitative impact on our results. That said, we want to acknowledge this shortcoming of our model and given this discussion, our model is likely to approximate mature firms best and not small startups. A model more suited to describing the economics of startups would be in Hartman-Glaser et al. (2019).

Last but not least, the formula describing the tradeoff along the joint upper boundary is little changed in the alternative setup. The full derivation closely follows that for the benchmark model and we therefore do not present it here. To summarize the differences, recall that the bound on compensation is now $p_w(c, w) + p_c(c, w) \geq -1$ and can reach a minimum of $p_w(\bar{c}(w), \bar{w}(c)) = -2$. Using this condition rather than $p_w(\bar{c}(w), \bar{w}(c)) = -1$ changes the slope along the joint upper boundary to be

$$\frac{dw}{dc} = -\frac{r\tau C}{2(\gamma - r)} < 0.$$ 

As we discuss in Section III our model is better suited to measuring relative and not absolute distortions. Therefore, a change to this alternative setup will not impact our steady state analysis of the changes in relative distortions.

**B. Data Appendix**

We use all industrial, standard format, consolidated accounts of firms in Compustat. We exclude firms without a NAICS code and in the utilities (22), financial (52-53), other (91), and public (92) industries. As is standard in the literature, we remove firms with missing or non-positive book assets ($at$) or sales ($sale$) and those with net property, plant, and equipment ($ppent$) of less than five million dollars. Our data sample starts in 1993, when compensation data from Execucomp becomes virtually complete, and ends in 2017. Following Gillan, Hartzell, Koch and Starks (2018) and McKeon (2015), we exclude observations where salary is available yet the item $tdc1$ is missing to minimize backfilling bias and define refinancing as common stock issuance greater than 5 percent of book assets. Because in the model state variables are defined over $K$ and in the data over assets $(C + K)$, we subtract cash from assets in the data to make variables comparable and define net assets as book assets less cash, $at - che$.

- **Cash Holdings** = cash (che) / net assets
- **Compensation** = (salary + bonus + LTIP + equity rewards) ($tdc1(t)$) / net assets ($t-1$)
- **Payout** = 1 if $dvc > 0$ or repurchases $> 0$; 0 otherwise
- **Free Cash Flow** = (EBITDA ($ebitda(t)$) - physical investment ($capx(t)$)) / net assets ($t-1$)
- **Investment** = (physical investment ($capx(t)$)) / net assets ($t-1$)
- **Preferred Issuance** = Use $\max(pstk(t) - pstkr(t-1), 0)$, $\max(pstk(t) - pstk(t-1), 0)$, or zero, in decreasing order of preference
- **Refinancing** = 1 if sale of common stock ($sstk$ less preferred issuance) / assets ($at$) $> 0.05$; 0 otherwise
- **Refinancing Size** = sale of common stock / net assets where refinancing = 1
- **Repurchases** = repurchases of common stock ($prstkc$ less preferred repurchases ($prstkpc$))
### Table I: Variable Definitions and Calibration

This table defines model variables and the values of the benchmark calibration discussed in Section III. All parameters are annualized.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Symbol</th>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital Stock</td>
<td>$K$</td>
<td>Agency Costs</td>
<td>$(\lambda_K, \lambda_C)$</td>
<td>(0.04, 0.09)</td>
</tr>
<tr>
<td>Cash Holdings</td>
<td>$C$ or $c = C/K$</td>
<td>Average Productivity</td>
<td>$\mu$</td>
<td>0.17</td>
</tr>
<tr>
<td>Managers’ Stake</td>
<td>$W$ or $w = W/K$</td>
<td>Volatility of Productivity</td>
<td>$\sigma$</td>
<td>0.35</td>
</tr>
<tr>
<td>Investors’ Value</td>
<td>$P(\cdot)$ or $p(\cdot) = P(\cdot)/K$</td>
<td>Depreciation</td>
<td>$\delta$</td>
<td>0.08</td>
</tr>
<tr>
<td>Investment</td>
<td>$I$ or $i = I/K$</td>
<td>Capital Adjustment Cost</td>
<td>$\theta$</td>
<td>6</td>
</tr>
<tr>
<td>Adjustment Cost</td>
<td>$G(\cdot)$ or $g(\cdot) = i + G(\cdot)/K$</td>
<td>Exogenous Expansion Rate</td>
<td>$z$</td>
<td>0.09</td>
</tr>
<tr>
<td>Cumulative Productivity</td>
<td>$A$</td>
<td>Recovery Rates</td>
<td>$(l_K, l_C)$</td>
<td>(0.4, 1)</td>
</tr>
<tr>
<td>Cumulative Free Cash Flow</td>
<td>$Y$</td>
<td>Interest Rate</td>
<td>$r$</td>
<td>0.040</td>
</tr>
<tr>
<td>Cumulative Payments</td>
<td>$U$</td>
<td>Management’s Discount Rate</td>
<td>$\gamma$</td>
<td>0.048</td>
</tr>
<tr>
<td>Cumulative Payouts</td>
<td>$D$</td>
<td>Refinancing Costs</td>
<td>$(\Phi, \phi)$</td>
<td>(0.005, 0.05)</td>
</tr>
<tr>
<td>Cumulative Refinancing</td>
<td>$F$</td>
<td>Corporate Tax Rate</td>
<td>$\tau_Y$</td>
<td>0.30</td>
</tr>
<tr>
<td>Cumulative Refinancing Costs</td>
<td>$X$</td>
<td>Penalty on Cash Holdings</td>
<td>$\tau_C$</td>
<td>0.20</td>
</tr>
<tr>
<td>Termination Time</td>
<td>$\tau$</td>
<td>Managers’ Relative Bargaining Power</td>
<td>$\psi$</td>
<td>0.5</td>
</tr>
<tr>
<td>Payment Boundary</td>
<td>$\bar{w}$ or $\bar{w}(c)$</td>
<td>Average Entrant Cash Holdings</td>
<td>$\mathbb{E}[c_0]$</td>
<td>0.15</td>
</tr>
<tr>
<td>Payout Boundary</td>
<td>$\bar{c}$ or $\bar{c}(w)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refinancing Size</td>
<td>$f$ or $f(w)$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This table reports averages and percentiles of several variables targeted in the data by the model's stationary distribution. The data annually cover the period from 1993 until 2017 and definitions are in Appendix B. Data variables are winsorized across all firm-years by 5 percent at the upper and lower tails except for refinancing size which is only winsorized at the upper tail. Percentiles are from the 25th, 50th, and 75th breakpoints. In the model, continuous variables are scaled cash holdings, $c = C/K$, compensation, $\gamma w = \gamma W/K$, investment $i = I/K$, and free cash flow $\mathbb{E}[dY]/K + \tau \gamma \delta$. Indicator variables are payout rate (47), entry/exit rate (44), refinancing rate (46), and refinancing size which is $f(w)$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Cash Holdings</td>
<td>21.8</td>
</tr>
<tr>
<td>Compensation</td>
<td>1.4</td>
</tr>
<tr>
<td>Investment</td>
<td>8.1</td>
</tr>
<tr>
<td>Free Cash Flow</td>
<td>4.0</td>
</tr>
<tr>
<td>Payout Rate</td>
<td>37.9</td>
</tr>
<tr>
<td>Entry/Exit Rate</td>
<td>1.2</td>
</tr>
<tr>
<td>Refinancing Rate</td>
<td>16.4</td>
</tr>
<tr>
<td>Refinancing Size</td>
<td>14.4</td>
</tr>
</tbody>
</table>
TABLE III: STEADY STATE ANALYSIS (ANNUAL)

This table reports averages under the stationary density from various calibrations of the model. Variables are scaled cash holdings, $c = C/K$, compensation, $\gamma_w = \gamma W/K$, investment $i = I/K$, and free cash flow $\mathbb{E}[dY]/K + \tau_Y \delta$. Indicator variables are payout rate (47), entry/exit rate (44), refinancing rate (46), and refinancing size which is $f(w)$. Agency and financial distortions are computed respectively in (48) and (49).

<table>
<thead>
<tr>
<th></th>
<th>Benchmark</th>
<th>$\mu = 0.18$</th>
<th>$\tau_Y = 0.21$</th>
<th>$\Phi = 0.015$</th>
<th>$\Phi = 0.015$</th>
<th>$\psi = 0.03$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Cash Holdings</td>
<td>21.8</td>
<td>25.1</td>
<td>23.3</td>
<td>24.9</td>
<td>25.2</td>
<td>25.0</td>
</tr>
<tr>
<td>Compensation</td>
<td>1.4</td>
<td>1.5</td>
<td>1.6</td>
<td>1.4</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>Investment</td>
<td>8.1</td>
<td>9.0</td>
<td>8.7</td>
<td>8.1</td>
<td>8.6</td>
<td>8.6</td>
</tr>
<tr>
<td>Free Cash Flow</td>
<td>4.0</td>
<td>4.4</td>
<td>4.7</td>
<td>4.1</td>
<td>4.7</td>
<td>3.9</td>
</tr>
<tr>
<td>Payout Rate</td>
<td>37.9</td>
<td>37.1</td>
<td>40.1</td>
<td>37.1</td>
<td>39.0</td>
<td>36.4</td>
</tr>
<tr>
<td>Entry/Exit Rate</td>
<td>1.2</td>
<td>0.8</td>
<td>0.1</td>
<td>1.8</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Refinancing Rate</td>
<td>16.4</td>
<td>13.8</td>
<td>18.0</td>
<td>13.8</td>
<td>15.9</td>
<td>14.1</td>
</tr>
<tr>
<td>Refinancing Size</td>
<td>14.4</td>
<td>18.5</td>
<td>17.5</td>
<td>19.5</td>
<td>20.7</td>
<td>16.9</td>
</tr>
</tbody>
</table>

Distortions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>$\tau_Y = 0.21$</th>
<th>$\lambda_C = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agency (a)</td>
<td>2.54</td>
<td>2.64</td>
<td>1.66</td>
<td>1.83</td>
</tr>
<tr>
<td>Financial (f)</td>
<td>0.27</td>
<td>0.25</td>
<td>0.33</td>
<td>0.35</td>
</tr>
<tr>
<td>Ratio (a/f)</td>
<td>9.52</td>
<td>10.71</td>
<td>5.10</td>
<td>6.28</td>
</tr>
<tr>
<td>Relative to (1)</td>
<td>1.00</td>
<td>1.13</td>
<td>0.54</td>
<td>0.64</td>
</tr>
</tbody>
</table>

47
This figure depicts the solution to the cash management problem with no agency conflict under the parameters tabulated in Table I with the exception of setting \((\mu, z) = (0.16, 0.02)\). The domain for both panels is scaled cash holdings, \(c = C/K \in [0, \bar{c}]\), where \(\bar{c}\) is the payout boundary. In Panel A, the blue curve is the firm’s enterprise value, \(p(c) - c\). The red dashed line depicts marginal financing costs, \(\Phi + \phi c\), and its tangency with enterprise value determines the refinancing size, \(f\). The stationary density is unscaled to the vertical axis and shown in gray. Panel B plots the investment rate and its decision equation.
This figure plots the solution to the agency problem with costless refinancing under the parameters tabulated in Table I with the exception of setting $(\mu, z) = (0.15, 0.02)$. The domain for both panels is managers’ stake (scaled continuation payoff), $w = W/K \in [0, \bar{w}]$, where $\bar{w}$ is management’s payment boundary. In Panel A, the blue curve is investors’ scaled value function. At the termination boundary, $w = 0$, investors recover $l_K$, and a new firm is drawn with initial payoff $w_0 = \psi \bar{w} + (1 - \psi)\bar{p}$, where $\bar{p} = \arg\max_w p(w)$. The stationary density is unscaled to the vertical axis and shown in gray. Panel B plots the investment rate and its decision equation.

**Figure 2: Agency Problem with Costless Refinancing**
This figure shows the accuracy of the complete model solution under the parameters tabulated in Table I. Panel A plots the first derivatives of investors’ scaled value function with respect to managers’ stake (scaled continuation payoff), \( w = W/K \), for percentiles of the marginal distribution of scaled cash holdings, \( c = C/K \). Panel B plots the first derivatives of investors’ scaled value function with respect to scaled cash holdings for percentiles of the marginal distribution of managers’ stake. Panel C plots the super contact condition for the payment boundary, the second derivative of \( p(c, \overline{w}(c)) \) with respect to \( w \) for each value of \( c \) and Panel D plots the super contact condition for the payout boundary, the second derivative of \( p(\overline{c}(w), w) \) with respect to \( c \) for each value of \( w \). The marginal densities of scaled cash holdings and managers’ stake are unscaled to the vertical axes and shown in gray.
This figure summarizes properties of the model solution under the parameters tabulated in Table I. Panel A plots investors’ scaled value function from above, bounded by the non-rectangular state space over managers’ stake, \( w = \frac{W}{K} \), and cash holdings, \( c = \frac{C}{K} \). The black dotted line is the size of refinancing conditional on managers’ stake, \( f(w) \). The red line is the payment boundary, \( \bar{w}(c) \), and the blue line the payout boundary, \( \bar{c}(w) \). These boundaries intersect at the joint upper boundary, \((\bar{c}(w), \bar{w}(c))\), marked by the black square. Panel B plots the stationary density. In both panels a brighter color represents a higher value.

**FIGURE 4: PROPERTIES OF MODEL SOLUTION**
This figure illustrates the calculation used to measure the agency distortion relative to second-best; an analogous calculation holds for the financial distortion. It plots the stationary density in green from above, bounded by the state space over managers’ stake, $w = W/K$, and scaled cash holdings, $c = C/K$. A darker green represents a higher value. The red line is the payment boundary, $\bar{w}(c)$, and the blue line the payout boundary, $\bar{c}(w)$. These boundaries intersect at the joint upper boundary, $(\bar{w}(c), \bar{c}(w))$, marked by the black square. The second-best frontier, the dash-dotted line, touches the joint upper boundary and has slope $\frac{\tau C}{\gamma - r}$. Next, the point $(c, w)$ marked by an O represents the density of firms at that point. The probability density of cash holdings over the next year point starting from O is normal with mean $E_t[dc]$ that is marked on the figure and standard deviation $(1 - \tau Y)\sigma$. This normal density is over the support denoted by the horizontal dotted line extending from O out to the second-best frontier, but is displayed with a tilt for the reader. The measurement for point $(c, w)$ is the expected value over the interval starting at $\bar{w}(w)$ and ending at the frontier $F^{-1}(w)$. 

**Figure 5: Illustration of Measurement of Distortions**
This figure plots the impulse response function of several firm policies of the same firm initialized at \((c, w)\) but contrasted across two economies that differ by fixed costs of refinance: \(\Phi = 0.5\%\) or \(1.5\%\). All variables are in percent. We initialize scaled cash holdings \(c = C/K\) near the average refinancing amount in the data, \(c_0 = 0.15\), and managers’ stake at \(w = (W/K)/\gamma = 0.013/\gamma\), where 1.3 percent is the average size of compensation in the data. Dynamics follow from the drifts of (38) and (39). Refinancing and payout statistics and agency and financial distortions are from (46), (47), (48) and (49), respectively.
This figure displays in Panels A and B the concavity of the value function required for $\beta \geq \lambda(c)/(\mu(1 - \tau_Y))$ to be the optimal solution and in Panels C and D the condition of full effort ($e_t = 1$) to be the optimal incentive strategy as we discuss in Appendix A. Panels A and B show $p_{ww}$ and the sum $\beta p_{ww}/2 + p_{cw}$ with respect to $w$ and $c$, respectively, across the 25th, 50th, and 75th percentiles of the marginal distribution of $c$ and $w$, respectively. Panels C and D plot the value of the inequality that must be positive to ensure that full effort is preferred to a policy in which agents shirk ($e_t = 0$). The upper and lower 10 percent of the marginal distribution of $w$ and $c$ are dotted lines and the intermediate 10-90 percent are solid lines. The black dashed lines mark zero.
This figure hypothetically illustrates the refinancing decision and the state space of the alternative model where managers are paid out of the firm’s cash holdings. The refinancing line is marked by the dashed line. The state space is the solid line.