A MODEL OF MARKET DISCIPLINE*

Colin Ward
University
of Minnesota

Chao Ying
Chinese University
of Hong Kong

17 December 2021

Abstract

We develop an equilibrium model where refinancing and managerial incentives are jointly determined to quantify external finance's influence on management's ex ante behavior. A greater fixed cost of refinancing lowers the cost of deferring compensation and, effectively, agency conflicts. The model novelly produces a transparent tradeoff between agency and financing distortions and, respectively, payout and compensation policy. Our analysis of US public firms predicts that the majority of a firm's free cash flow is directed towards mitigating agency conflicts rather than financial frictions and suggests that introducing a tax on refinancing can reduce the relative severity of agency costs.

*Ward: Carlson School of Management, University of Minnesota; cward@umn.edu. Ying: Department of Finance, CUHK Business School; chaoying@cuhk.edu.hk. We thank Hengjie Ai, Patrick Bolton, Yasser Boualam, Barney Hartman-Glaser, Zhiguo He, Burton Hollifield, Jonathan Payne, Martin Szydlowski, Han Xiao and participants at EWMES, MFA, 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 managers' incentives with investors' objectives. 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 through which channel does it work? More broadly, how do agency conflicts and financial frictions jointly affect firm behavior and can they be altered with policy? 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 begin 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 condition its choice on the firm's performance history. Managers understand this and, therefore, it affects their behavior and decisions today. We call this equilibrium effect *market discipline*.

Our paper makes two contributions. First, by formalizing the interaction between financial and agency distortions, 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. Our model shows us that raising the fixed cost of refinancing has the effect of reducing the cost of deferring managers' compensation and, effectively, agency conflicts. A tradeoff thus exists between financing frictions and agency conflicts.

Second, we derive a novel formula that connects this tradeoff to the allocation of free cash flow towards investors or managers:

$$\frac{\text{Size of Agency Conflict}}{\text{Size of Financial Friction}} \propto \frac{\text{Free Cash Flow Paid to Investors}}{\text{Free Cash Flow Paid to Managers}}.$$

For intuition, consider a firm with scarce cash holdings. Here the marginal cost of restricting payouts to investors is low as the likelihood of costly refinancing is large. Payments made to managers, during which agency conflicts are muted, *must* then correspond to a financial distortion that is present in the firm.

Fixing cash holdings and varying payments to managers measures the financial friction much like varying supply identifies a demand elasticity. More generally, 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 introduce the key economic tradeoffs of the model of Section I to better understand our headline results. To separate their interests from investors', managers are relatively impatient and can possibly consume private benefits. Their private consumption comes at the expense of their effort that enhances asset efficiency and reduces the likelihood of costly liquidation. Investors therefore write a contract that can be terminated at any time and that uses deferred compensation to incentivize managers to work.

External financing is costly to investors and has a fixed and variable component. Together, these costs imply that cash will be raised intermittently and in finite amounts and is the preferred source to finance investment. But because cash earns a smaller return than what investors could earn elsewhere, its accumulation will be limited.

While its costs are predetermined, the quantity of refinancing is endogenously determined within the incentive contract. Because the contract is history-dependent, refinancing depends on the historical performance of the firm that, in turn, depends on the commitment to effort chosen by managers. The discipline created from refinancing thus acts as the bridge through which managers internalize investors' preferences, altering their decisions today.

We show in Section III that a more stringent cost of refinancing, for example, translates into a stricter effort policy that management choose to uphold to receive their (promised) deferred compensation. In effect, a greater cost of external finance has lessened the cost to investors of deferring compensation to managers—financial frictions trade off with agency conflicts.

Of course, the firm's investors strive to minimize the harmful effects of agency and financial distortions on firm value. Promising an additional dollar of deferred compensation and holding an additional dollar of cash will, respectively, attenuate these distortions, but only up to a point. The marginal benefits of an additional dollar falls yet inefficiencies remain—managers' impatience and cash's inferior rate of return.

At the threshold where the marginal benefits equal the marginal costs due to these inefficiencies, we characterize explicitly the tradeoff between financial frictions and agency conflicts. We show that beyond this threshold, free cash flow is allocated freely to investors and managers according to the ratio of agency to financial distortions. The insight is therefore that the allocation of free cash flow is informative about the underlying frictions. This closes our understanding of the origins of the novel formula that we derive in Section III

Our measurement of frictions has advantages over those in classic agency or costly external financing models. First, classic models measure frictions as deviations from first-best, which may be neither attainable practically nor serve as a reasonable benchmark. In our model, we measure deviations from second-best: for example, the magnitude of financial frictions controlling for the degree of agency conflicts. Second, they often make comparisons implicitly based on market values that are influenced by hard-to-measure discount rates. An appealing feature of our theoretically-grounded approach is that payouts to investors and payments to managers can be observed with readily available accounting data.

More broadly, a burgeoning literature studies the impact of financing frictions on the macroeconomy. Our paper shows and we argue that agency conflicts are a serious distortion deserving commensurate study. Indeed, our calibrated model of Section IV suggests that they distort the decisions of firms more than do financial frictions in the United States, at least secularly over frequencies lower than the business cycle's. This is because while cash can simply be accumulated to minimize instances of costly external finance, it is double-edged as it exacerbates the alignment of managerial incentives.

Our calibration 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 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. This 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 100 basis points, which could be implemented with a small tax on the event of refinancing. We find that a lower corporate tax helps offset the cost of a refinancing tax and further allows them to allocate additional cash flow towards mitigating agency conflicts. In effect, raising the fixed cost of refinance lowers the cost of deferring compensation.

Literature

Our paper analyzes the capital market implications of dynamic agency.[1] 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 the agent's 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. Similar to this paper which takes the structure of adjustment costs as given, we take the structure of refinancing costs of given and analyze the policies of the optimal contract.

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.[2] Bolton, Chen and Wang (2011) show that the marginal value of cash affects investment, external financing,

[1] A partial list is Albuquerque and Hopenhayn (2004), Quadrini (2004), Dow, Gorton and Krishnamurthy (2005), Clementi and Hopenhayn (2006), DeMarzo and Fishman (2007), Biais, Mariotti, Plantin and Rochet (2007), Ai and Li (2015), Morellec, Nikolov and Schürhoff (forthcoming), Boualam (2019), Ward (2019), and Tong and Ying (2019).

[2] A short list is Gomes (2001), Whited and Wu (2006), Hennessey and Whited (2007), Riddick and Whited (2009), Rampini and Viswanathan (2010), Nikolov and Whited (2014), Milbradt and Oehmke (2015), and Belo, Lin and Yang (2018).

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.[3] 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, Kalemli-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 an explanation towards the mysterious decline in US allocative efficiency (Bils, Klenow and Ruane (2020)).

I. ECONOMIC ENVIRONMENT

The environment is a standard neoclassical model of investment augmented with cash holdings, costly external finance, and agency conflicts. The economy's equilibrium stationary distribution is jointly determined with conditions of firm exit and entry. It is minimalist in the following sense: if only costly refinancing 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 the ability to finance internally, moreover, makes this costly event meaningful.

The economy consists of investors and managers, who are hired by investors to run firms. 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 $\gamma > r$ (see DeMarzo and Sannikov (2006) for details). They have no initial wealth and limited liability.

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 close the setup within the contracting environment.

[3] An incomplete list is King and Levine (1993), Rajan and Zingales (1998), Buera, Kaboski and Shin (2011), Asker, Collard-Wexler and de Loecker (2014), Haltiwanger, Kulick and Syverson (2018).

A. TECHNOLOGY, FREE CASH FLOW, AND AGENCY CONFLICT

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 $G(I, K)$ and we write the total cost of investment as $I + G(I, K)$. We assume that adjustment costs, $G(I, K)$, are homogeneous of degree one in its arguments. The total cost of investment is thus $I + G(I, K) = Kg(i)$, where $i = I/K$ is the investment rate and $g(i)$ is an increasing and convex function.

Free cash flow is the cash flow available to distribute to investors net of investment expenses and taxes at rate τ_Y. After optimally choosing and compensating freely adjustable labor (non-management), free cash flow is determined by a constant returns to scale technology:[4]

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

In contrast to the neoclassical framework, 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 with 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 σ. Shirking gives managers private benefits at rate λdt per unit of capital, an assumption validated in the long literature on empire building whereby managerial rents grow with

[4]For a given capital stock and with freely adjustable labor, L, the firm solves the static problem $\max_L (1 - \tau_Y)(a_t K_t^{\alpha} L_t^{1-\alpha} - w_L L_t)$, 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 α.

firm size and consistent with the thesis of Jensen (1986). The parameter $0 \leq \lambda \leq 1$ measures the severity of the agency problem.

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) gives a broad definition of a firm's resources as "the aggregate purchasing power available to management for strategic purposes during any given planning period." To narrow the distance between model and data, however, we restrict them to be only cash holdings.

Cash held at time t is denoted by C_t and its 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 Φ and a marginal cost ϕ.[5] 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 assume both costs scale with capital as informational costs or the effects of dilution are likely to be proportional to firm size (Bolton et al. (2011)). We denote the cumulative costs of external financing up until time t by X_t and its

[5] Similar to DeMarzo et al. (2012) who solve a dynamic contracting problem given the structure of adjustment costs, we solve a dynamic contracting problem given the structure of refinancing costs.

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 addition to refinancing, investors can choose to let the firm fail and liquidate it. To distinguish this choice, we now describe the contract between investors and managers.

C. Contracting Environment and Termination

Investors write a contract that gives managers incentive to exert effort and that can be terminated at any time. If termination occurs at time t, investors retrieve the firm's cash and receive a fraction of capital, $0 < l_K \leq 1$, altogether recovering $l_K K_t + C_t$. Managers obtain their outside option, normalized to zero.

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, τ, all of which depend on the entire history of productivity A_t. Limited liability requires U to be nondecreasing. We let $\mathcal{C} = (I, F, D, U, \tau)$ represent the contract.

Given the contract, management chooses an effort process to solve

$$W(\mathcal{C}) = \max_{\{e_t \in \{0,1\}: 0 \leq t < \tau\}} \mathbb{E}^e \left[\int_0^\tau e^{-\gamma t} \left(dU_t + \lambda K_t (1 - e_t) dt \right) \right], \tag{5}$$

where $\mathbb{E}^e[\cdot]$ 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 private benefits when taking action $e_t = 0$.

Upon contract initiation, 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

$$P(K_0, C_0, W_0) = \max_{\mathcal{C}} \mathbb{E}\left[\int_0^\tau e^{-rt}(dD_t - dF_t - dX_t - dU_t) + e^{-r\tau}\left(l_K K_\tau + l_C C_\tau\right)\right]$$
$$\text{s.t. } \mathcal{C} \text{ is incentive compatible and } W(\mathcal{C}) = W_0. \tag{6}$$

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.[6]

Management's initial payoff W_0 is determined by their relative bargaining power. 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, $W_0^I \equiv \text{argmax}_{W\geq 0} P(K_0, C_0, W)$. More generally, we locate an interior to the 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 and Market Discipline

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

$$W_t(\mathcal{C}) = \mathbb{E}_t\left[\int_t^\tau e^{-\gamma(s-t)} dU_s\right]. \tag{7}$$

Note the collapse of the expectation operator $\mathbb{E}^e[\cdot]$ 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 (Zwiebel (1996)) and formalizes the notion of costly refinancing 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)). The optimal contract compensates management for their time preference on average: $\mathbb{E}_t\left[dW_t + dU_t\right] = \gamma W_t dt$. To maintain incentive com-

[6]In Appendix A we discuss an alternative setup in which payments to managers dU are subtracted from dC in (4). Thus in our model we consider polices where owners pay managers directly and segregate the firm's cash holdings from compensation. This clarifies the analysis and allows a separation of compensation structure from internal financing policy. A consequence of this assumption is that we are implicitly assuming that firms do not refinance to pay managers current payments, like bonuses.

patibility, management's compensation must remain sufficiently sensitive to the firm's free cash flow, dY_t. Sannikov (2008) shows this sensitivity, β_t, is given by the martingale representation theorem:

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

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

$$\beta_t = \frac{\lambda}{(1-\tau_Y)\mu} \text{ for all } t. \qquad (9)$$

Intuitively, the optimal sensitivity is a ratio of the return to shirking λ to the return to working $(1-\tau_Y)\mu$. For a given beneficial shock to free cash flow ($dZ_t > 0$), a larger firm size is associated with a greater increase in managers' compensation, consistent with Edmans, Gabaix and Landier (2008).

Also important is that the shocks to agents' continuation utility are driven by productivity shocks and not by the event of refinancing directly. This assumption is how we choose to model market discipline as an indirect force that invisibly guides managers' behavior. More specifically, it implies that for a given set of state variables and the parameters $(\gamma, \lambda\sigma/\mu)$, the *instantaneous* distribution of 8 across differently calibrated economies would be identical. Firm value across different calibrations, of course, would differ, and consequently so too would the path of investment and capital.

II. Model Solution

Here we describe some properties of the solution to (6). The first-best solution is in Appendix A. 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.

Our model has the homogeneity property that 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:

$$rdt = \max_{i} \mathbb{E}_t \left[\frac{d(Kp(c, w))}{Kp(c, w)} \right], \tag{10}$$

subject to the incentive compatibility constraint in (9). This equation's solution is jointly determined with the boundaries that determine refinancing, payouts to investors, payments to managers, and contract termination, which we now discuss.

A. Cash Management Policy

The convex adjustment cost implies that investment will be smooth; therefore, the firm can achieve any investment rate so long as cash holdings are positive. There is thus no incentive to refinance early. Consequently, firms economize fixed refinancing costs by issuing equity intermittently, and will do so only when cash holdings fall to zero.

When refinancing, the firm chooses and receives a total issue amount of $f(w) > 0$ per unit of capital. Because firm value is continuous before and after issuance, value matching at the refinancing boundary $c = 0$ holds:

$$p(0, w) = p(f(w), w) - \Phi - (1 + \phi)f(w), \text{ for each } w. \tag{11}$$

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

$$p_c(f(w), w) = 1 + \phi, \text{ for each } w. \tag{12}$$

Conversely, because holding cash in the firm is penalized at rate τ_C, the firm will distribute it to investors when abundant. Formally, abundance takes the form of an endogenous payout boundary, which we denote by $\overline{c}(w)$. For $c > \overline{c}(w)$ firm value must be equal pre- and post-distribution, implying the equation $p(c, w) = p(\overline{c}(w), w) + (c - \overline{c}(w))$. Because this equation holds continuously, the limit $c \to \overline{c}(w)$ is summarized by the derivative

$$p_c(\overline{c}(w), w) = 1, \text{ for each } w. \tag{13}$$

Intuitively, at $\overline{c}(w)$ 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

$$p_{cc}(\overline{c}(w), w) = 0, \text{ for each } w. \tag{14}$$

B. OPTIMAL CONTRACTING POLICY

If the contract is terminated, investors receive the liquidation value of the firm. Termination is thus costly and this precludes investors from choosing it early. Nevertheless, management will be terminated once their continuation utility hits their outside option (set to zero) because otherwise they would immediately consume private benefits. Hence, investors' liquidation payoff at this termination boundary is

$$p(c, 0) = l_K + c, \text{ for each } c. \tag{15}$$

Thus, regardless of the level of cash the firm will be liquidated once $w = 0$.

Next, because investors can always compensate management with cash from their pockets, it will cost investors at most one dollar to promise another dollar of deferred compensation to managers: $p_w(c, w) \geq -1$. This implies that $p(c, w)$ is weakly increasing in w and that it is optimal to grow w quickly when its value is low. Hence, there is a benefit to deferring compensation—the reduction in the probability of costly termination—and this is achieved by setting incremental current payments dU/K in (8) to zero. As management is relatively impatient, however, at some point they will need to receive current payments. The optimum is located where the marginal benefit and marginal cost are equalized and, formally, takes the form of a payment boundary that is the threshold where investors are indifferent between promising another dollar to managers and paying them immediately

$$p_w(c, \overline{w}(c)) = -1, \text{ for each } c. \tag{16}$$

Finally, because this threshold is determined optimally it must satisfy the super contact condition

$$p_{ww}(c, \overline{w}(c)) = 0, \text{ for each } c. \tag{17}$$

C. CROSS-PARTIAL BOUNDARIES

We now study the boundaries implied by cross-partial derivatives. Recall that the upper thresholds $\overline{c}(w)$ and $\overline{w}(c)$ are pinned down by optimality conditions, decisions which need to hold with respect to both states. Economically, this means that along a segment of $\overline{c}(w)$, where investors receive payouts, it is optimal to not give managers current payments; and similarly when managers receive payments along $\overline{w}(c)$ that investors will not earn payouts. Of course, there could be segments on which $\overline{w}(c)$ and $\overline{c}(w)$ 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 $\overline{c}(w)$ achieves super contact we first differentiate $p(c, w) = p(\overline{c}(w), w) + (c - \overline{c}(w))$ with respect to w which gives

$$p_w(c, w) = \frac{\partial p(\overline{c}(w), w)}{\partial w} - \frac{\partial \overline{c}(w)}{\partial w}, \text{ for each } c \geq \overline{c}(w). \tag{18}$$

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

$$p_{wc}(\overline{c}(w), w) = 0, \text{ for each } w. \tag{19}$$

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

$$p_{cw}(c, \overline{w}(c)) = 0, \text{ for each } c. \tag{20}$$

Finally, it is evident from equations (19) and (20) that

$$p_{cw}(\overline{c}(w), \overline{w}(c)) = 0, \text{ for every } c \text{ and } w. \tag{21}$$

D. INTERIOR REGION, HJB EQUATION, AND INVESTMENT

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. Within the boundaries, the system evolves according to the dynamics of the endogenous state variables (1), (4), and (8) that allows us to write, by Ito's lemma,

$$dc_t = ((1-\tau_Y)\mu - g(i_t) + \tau_Y\delta + [r(1-\tau_C) - (i_t-\delta)]c_t)dt + \sigma(1-\tau_Y)dZ_t \tag{22}$$

and $$dw_t = (\gamma - (i_t - \delta))w_t dt + \lambda\frac{\sigma}{\mu}dZ_t. \tag{23}$$

Both cash holdings 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.

In what follows we assume that $\beta p_{ww}/2 + p_{cw}$ and p_{ww} are nonpositive, so that it is indeed optimal for the contract to always prescribes effort. The solution to (10), then, can be represented by the partial differential equation

$$\begin{aligned} rp(c,w) = \max_i\, & p(c,w)(i-\delta) + p_c(c,w)\left((1-\tau_Y)\mu - g(i) + \delta\tau_Y + [r(1-\tau_C)-(i-\delta)]c\right) \\ & + p_w(c,w)\left((\gamma-(i-\delta))w\right) + \frac{1}{2}p_{cc}(c,w)(\sigma(1-\tau_Y))^2 \\ & + \frac{1}{2}p_{ww}(c,w)\left(\frac{\sigma}{\mu}\lambda\right)^2 + p_{cw}(c,w)\frac{\sigma^2(1-\tau_Y)}{\mu}\lambda. \end{aligned} \tag{24}$$

subject to the resource boundaries of (11), (12), (13), and (14) that locate the position and ensure the optimality of the refinancing and payout decisions, $f(w)$ and $\overline{c}(w)$; the incentive boundaries in (15), (16), and (17) that determine termination and the payment threshold to management, $\overline{w}(c)$; as well as the mixed boundaries given by (19), (20), and (21). We detail our computational method to solve this problem in Appendix B. An important novelty in our computational solution relative to the problem studied by Achdou, Han, Lasry, Lions and Moll (forthcoming) is that in our setup (24) is solved jointly with the shape of the state space.

The drifts of (22) and (23) are impacted by the optimal choice of investment:

$$g'(i) = \frac{p(c,w) - p_w(c,w)w}{p_c(c,w)} - c \tag{25}$$

In a frictionless world, the marginal adjustment cost $g'(i)$ would be equated with marginal q, which under homogeneity would equal average Q (Abel (2018)). With costly external financing, the investment decision now accounts for financial frictions. In particular, a greater marginal cost of financing—that from optimality equals the marginal value of cash—lowers investment, all else equal. Furthermore in the presence of agency conflicts, investors now internalize the effect of investment on managers' incentives. For a given $w = W/K$, an increase in capital reduces management's effective claim on the firm and exacerbates the agency friction.

The decision deepens the link between cash holdings, compensation, and investment and is broadly consistent with facts established in the data. 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.

E. Distortions and The Joint Upper Boundary

The problem that the firm's investors face is of choosing investment to maximize value while navigating productivity shocks in the presence of confounding financial and agency factors. Ideally, investors would locate the firm where both confounders and their detrimental impact on value are minimized, where $p_c(\overline{c}(w), w) = 1$ and $p_w(c, \overline{w}(c)) = -1$.

A remarkable result is that in spite of the problem's complexity, the tradeoff faced by the firm's investors along the joint upper boundary is lucid:

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

$$\frac{dw}{dc} = -\frac{r\tau_C}{\gamma - r} < 0 \tag{26}$$

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 ($p_c(\cdot) = 1$ and $p_w(\cdot) = -1$), the lingering sources of inefficiency, the cash tax penalty $r\tau_C$ and managers' impatience $\gamma - r$, survive and determine the optimal tradeoff faced by firms.

More specifically, if investors decide to hold an additional amount dc of cash and refrain from paying themselves a payout, they will bring managers' payment boundary inwards by $\frac{r\tau_C}{\gamma - r}dc$, making it more likely management will receive immediate compensation. The proposition therefore 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 deferred compensation's marginal value $p_w(\cdot) > -1$ as it is becomes more valuable to avoid costly termination. Therefore as either c or w decrease, so do the marginal values that change the tradeoff faced along the boundary.

This intuition implies that 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.[7]

F. 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.

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

$$\frac{\partial h(c, w, t)}{\partial t} = \varphi(c, w)m + \mathcal{A}^* h(c, w, t), \tag{27}$$

where $\mathcal{A}^* h(c, w, t)$ is the adjoint of the infinitesimal generator of the bivariate diffusion

[7] In our setup, the tradeoff is exactly linear at the joint upper boundary and a different setup, perhaps one that has 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.

process (dc_t, dw_t).[8] 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: $\int_0^{\bar{c}(w)} \int_0^{\overline{w}(c)} h(c, w) dwdc = 1$. After this normalization notice the left side of (27), twice integrated, is zero and we can then rearrange it for the stationary entry rate, which by construction equals the exiting mass of firms,

$$m = -\int_0^{\bar{c}(w)} \int_0^{\overline{w}(c)} \mathcal{A}^* h(c, w) dwdc. \tag{28}$$

When a firm's contract is terminated, an entrant's cash holdings is drawn from a distribution with positive support. Its initial continuation payoff, w_0, moreover, is determined by the bargaining power of agents and investors.

III. Analysis of Market Discipline

In this section we analyze the effect of market discipline on our model: specifically, the effect of the fixed cost of refinance on managers' incentives. We do this by changing the parameter Φ and studying the impact on three of the firm's decisions: refinancing, investment, and the thresholds of payouts and payments.

In spirit with much of the dynamic contracting literature, see for example DeMarzo et al. (2012), our parameters here are chosen for illustration. In Appendix A we show that this calibration satisfies both the boundary conditions of the model as well as the full effort condition. In Section IV we later calibrate our model to salient moments of the data to assess the model's quantitative predictions.

A. Enterprise Value and Refinancing

We first study the effect of the fixed refinancing cost on firm value and the refinancing decision. Our notion of firm value is *enterprise value* $P(K, C, W) + W - C$, the total

[8] Specifically, $\mathcal{A}^* h(c, w, t)dt = -\mathbb{E}_t[dc] h_c(c, w, t) - \mathbb{E}_t[dw] h_w(c, w, t) + \frac{1}{2}\mathbb{E}_t[(dc)^2] h_{cc}(c, w, t) + \frac{1}{2}\mathbb{E}_t[(dw)^2] h_{ww}(c, w, t) + \mathbb{E}_t[(dcdw)] h_{cw}(c, w, t)$.

value of marketable claims net of cash. The value per unit of capital is

$$\frac{P(K,C,W)+W-C}{K} = p(c,w) + w - c. \tag{29}$$

Enterprise value is allocated across investors $p(c,w)$ and managers w. Accumulating cash holdings enhances enterprise value because $p_c(\cdot) \geq 1$. Raising managers' continuation utility w also increases the value of marketable claims held by both investors and managers, $p(c,w)+w$ because $p_w(\cdot) \geq -1$.

Recall that the firm can be refinanced or liquidated at any time. Both events become more probable following a sequence of adverse shocks, and therefore they are endogenously linked in our model. This is in contrast to Bolton et al. (2011) where liquidation or refinancing is determined by a choice of parameters.

In our model, it is optimal to refinance only when cash holdings reach zero, regardless of the level of managers' scaled continuation utility w. Nevertheless, investors could always choose not to refinance and terminate the contract, destroying managers' deferred compensation and leaving them with their outside option (zero). Thus, despite refinancing costs being solely borne by investors, the fact that the contract can be terminated at any time places discipline on managers in the form of a punishment that is determined off the equilibrium path.

To see this, we rearrange the value matching condition of refinancing seen in (11) to

$$(p(f(w),w) + w - f(w)) - (p(0,w) + w) = \Phi + \phi f(w). \tag{30}$$

Thus the marginal gain in enterprise value from refinancing, $(p(f(w),w) + w - f(w)) - (p(0,w)+w)$, is equated to its marginal cost, $\Phi+\phi f(w)$. We depict this decision in Figure 1 for a fixed level of w. Basic economic intuition shows that if the fixed cost of refinancing is now higher, then it is optimal to issue more upon refinancing and $f(w)$ increases. Moreover, the marginal value of cash has also risen and investors will decide to raise the payout boundary, $\overline{c}(w)$.

What is new is the effect on enterprise value, which *grows* with Φ. We emphasize that this result is in stark contrast to Bolton et al. (2011). In their model, a greater fixed cost of refinancing mechanically lowers enterprise value as it is a reflection of the present value of payouts net of refinancing costs. In our model, a larger cost places more discipline on managers' behavior via the threat of contract termination. This highlights the key mechanism in our model of market discipline.

B. Marginal Enterprise Value and Investment

The definition of enterprise value can be used to define the marginal benefit of enterprise value with respect to another dollar of cash or deferred compensation

$$\frac{\partial(p(c,w)+w-c)}{\partial c} = p_c(c,w) - 1 \geq 0 \quad \text{and} \quad \frac{\partial(p(c,w)+w-c)}{\partial w} = p_w(c,w) + 1 \geq 0. \tag{31}$$

Both of these margins are monotonically decreasing in their arguments, bounded below by zero, and encapsulate the current tradeoff each state variable presents to the firm's stakeholders.

For insight into the investment decision in (25), we rearrange to form

$$p(c,w) + w - c = p_c(c,w)g'(i) + (p_c(c,w) - 1)c + (p_w(c,w) + 1)w. \tag{32}$$

The left side is the gain in enterprise value that an additional unit of capital creates. The right side decomposes three marginal costs. First, the traditional marginal adjustment cost $g'(i)$ is now weighted by the marginal value of cash, as investment is financed internally. Second, as cash is paid out the firm becomes more financially constrained, a cost reflected in $p_c(c,w) - 1$. Third, investment shrinks the share of marketable claims held by managers $w/(p(c,w)+w)$, worsening the agency conflict as measured by $p_w(c,w) + 1$.

Figure 2 plots this decomposition for a low and high level of Φ. We see that the increase in enterprise value is associated with greater investment, echoing Tobin's intuition. The decomposition of costs shows that raising the fixed cost of refinance primarily influences the marginal value of deferred compensation rather than of cash, especially at small cash holdings. Thus, the net marginal benefit of deferred compensation has grown: the benefit of reducing the probability of termination is now larger or management has become, in effect, more patient and this reduces the cost of deferring compensation. In effect, the agency conflict has been attenuated, leading to the gain in enterprise value.

C. Boundaries and Refinancing Amount

We now turn to analyzing the shapes of the payout and payment boundaries as well as the quantity of refinancing $f(w)$ before discussing how a change in the fixed refinancing cost affects these objects. Figure 3 displays these decisions under the low and high Φ parameterizations.

For either parameterization, the boundary touches the zero axis for both cash holdings and managers' stake. Starting from either zero axis and remaining on a curve while moving in to the interior, the boundary curve eventually reaches the joint upper boundary, marked by the black squares and the associated tangent lines implied by the Proposition.

C.1. Management's Payment Boundary, $\overline{w}(c)$

The payment boundary is a policy function that investors will choose to raise or lower the threshold depending on the level of cash. The rise of cash from zero coincides with a decline in the likelihood of costly refinancing that leads to a gain in enterprise value. Now that investors are faced with a greater dollar loss in the event of termination, it is efficient to raise the payment boundary $\overline{w}(c)$ outwards and reduce the probability of liquidation.

At some point during the climb in cash holdings, however, the payment boundary begins to come back in. The continuing decline in the marginal value of cash coincides with a growing liquidation value. Thus cash holdings eventually begin to lower the cost of termination and consequently the payment boundary.

We emphasize that our model predicts a novel motive for retaining cash in the firm. Traditional thinking places precautionary savings motives at the heart of the cash holding decision. In our model, the stock of cash holdings *per se* is immediately free from the impact of agency. Cash acts as insurance for investors against the early termination of the contract with managers.

C.2. Investors' Payout Boundary, $\overline{c}(w)$

The payout boundary is a choice dependent on management's stake. As w grows from a low value, the agency conflict's deleterious effect on firm value weakens. This, in turn, raises the marginal value of capital, making it optimal to keep more cash in the firm to finance investment and to delay payouts by raising $\overline{c}(w)$.

Further increases in w beyond $p_w(\cdot) = 0$, however, result in a wealth transfer from investors to managers. Enterprise value still grows with w, but an ever increasing share of it is earmarked for managers. Managers thus have a keen interest in maintaining the level of continuation utility implied by their contract and will be agreeable in acting in investors' interest to preserve it.

Therefore, management begin to internalize refinancing's effect on on firm value. They realize that a large issue imposes an unnecessarily high cost on the firm's investors (and they could face a unexpectedly large loss if the contract is terminated off the equilibrium path). In response, managers work more voluntarily towards generating free cash flow, reducing the cost of financing constraints, and correspondingly lowering the payout boundary to investors. This internalization is reminiscent of Zwiebel (1996), who argues that managers voluntarily set debt to restrict themselves. Market discipline, even though present only ex post, acts as an ex ante constraint on management.

C.3. Refinancing Size, $f(w)$, and the Impact of Changing Φ

The refinancing curve $f(w)$ 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.

Refinancing initially increases with w. It is natural to believe that investors would likely provide more funds to a firm with a smaller agency conflict. While it initially rises rapidly away from $w = 0$, $f(w)$ later plateaus and eventually falls. As we just discussed, when management has a large enough stake in the firm and their incentives are well-aligned with investors', it becomes optimal to lower the size of refinancing, which is costly. The exact shape of $f(w)$ reflects these tradeoffs.

Finally, raising the fixed refinancing cost Φ amplifies the forces of market discipline. The payout boundary intuitively shifts outwards to reduce the frequency of refinance and, now that firm value is higher, it is efficient to raise the management's payment boundary.

The length of the interval in which the contract is terminated $c \in [0, \overline{c}(0)]$, however, falls. Because managers' incentives are better aligned with investors', optimality reduces the minimum level of cash held by the firm in which management could possibly be terminated.

D. MEASUREMENT OF DISTORTIONS

Traditional approaches to measuring the distortions of costly external finance and agency conflicts typically compare 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, a value 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 defines the slope of the joint upper boundary and implies the existence of a *second-best frontier* that is tangent to this boundary. Its existence means that we can use the model to measure distortions relative to this frontier. An advantage is that the measurements are based on quantities, which are readily observable and relatively accurate, 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 in reference to the payment boundary. We denote the frontier as a function of state by $\mathcal{F}(c)$ and $\mathcal{F}^{-1}(w)$ and in Appendix A we derive the formal definitions:

$$\text{Agency Distortion: } \mathbb{E}\left[\Delta c \middle| \overline{c}(w) < c + \Delta c < \mathcal{F}^{-1}(w)\right], \tag{33}$$

$$\text{Financial Distortion: } \mathbb{E}\left[\Delta w | \overline{w}(c) < w + \Delta w < \mathcal{F}(c)\right]. \tag{34}$$

These calculations are based on the annualized drifts and volatilities of (22) and (23) 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 $\overline{c}(w)$? Compare a firm at the joint upper boundary $(\overline{c}(w), \overline{w}(c))$ to one on $(\overline{c}(w), w)$ where $w < \overline{w}(c)$. For this firm, the financial friction is due only to the holding cash penalty, $r\tau_C$, since $p_c(\overline{c}(w), w) = 1$, so the only distortion from the second-best frontier can be attributed to agency. The idea is akin to using a supply shock to identify a demand elasticity and implies analogously that financial frictions can be inferred from patterns in management's compensation.

Put differently, payouts which occur between $\overline{c}(w)$ and $\mathcal{F}^{-1}(w)$ are *required* by investors to compensate them for holding a firm with an acute agency friction. Analogously, current payments which occur between $\overline{w}(c)$ and $\mathcal{F}(c)$ 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.

In the top two panels of Figure 4 we validate the relationship between our measurements and the marginal enterprise value with respect to both states. The relationship is monotone and nonlinear. Therefore, greater rates of payouts or compensation measure the underlying unobserved value distortions attributed to agency conflicts and financial frictions.

Panel C panel illustrates an exercise in the measurement of 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 (22) alone (recall our distortions' definitions are orthogonalized), we can evaluate the probability that this firm will receive a shock and cross the payout boundary $\overline{c}(w)$. For a given w, we truncate the likelihood of shock realizations beyond $\mathcal{F}^{-1}(w)$ as they surpass second-best outcomes. Since the shock changes cash holdings or compensation, we are measuring quantities.

To summarize, 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, and an alternative model would lead to different estimates.

In Hsieh and Klenow (2009), they control for this shortcoming by comparing the same model across countries and evaluating distortions relative to a benchmark economy. In our setup, we use a different, yet valid approach that computes 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 less sensitive to model choice.

IV. Calibration and Quantitative Analysis

Having studied the qualitative aspects of the model, we now calibrate it and discuss in detail the solution's quantitative properties.

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, a pertinent sample in which to study agency conflicts, 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 $\tau_Y = 30$ 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

$$g(i) = i + \frac{\theta}{2}(i - \delta - z)^2, \tag{35}$$

where θ 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 θ 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.

A.1. Internal Calibration: Averages

We internally calibrate our remaining six parameters $(\mu, \sigma, \gamma, \psi)$ 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, μ, 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 γ) 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 γ, 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 $\gamma = 0.048$.

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 (28). We target this rate with $\lambda = 0.06$, which influences the variation in management's continuation utility and the probability of contract termination, and $\sigma = 35$ percent to match the two percent average default rates of public firms over 1993 to 2017 (Boualam, Gomes and Ward (2020)). Thus our agency and volatility parameters are 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 σ, 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.

Consistent with this, the model calculation comes is motivated by the following 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, Δz, 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 $\mu_c(c, w)$, σ_c, $\mu_w(c, w)$, and $\sigma_w(c)$ are the annualized drifts and volatilities of (22) and (23). Given these moves, we calculate

$$\text{Refinancing Rate: } \mathbb{E}[\mathbf{1}\{\Delta c + c < 0\}|c, w] = \mathcal{N}\left(-\frac{\mu_c(c, w) + c}{\sigma_c}\right), \tag{36}$$

$$\begin{aligned}\text{Payout Rate: } &\mathbb{E}[\mathbf{1}\{\bar{c}(w + \Delta w) < \Delta c + c\}|c, w] \\ &= \mathbb{E}\left[\mathbf{1}\{\Delta z > (\bar{c}(w + \mu_w(c, w) + \sigma_w(c)\Delta z) - c - \mu_c(c, w))/\sigma_c\}\right],\end{aligned} \tag{37}$$

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 $\mathbb{E}[f(w)|w]$ 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 repurchases and special one-time dividends, 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. Quantitative Analysis

With all objects defined and the model solved, 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.

B.1. 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.

We find agency frictions to be 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 Farre-Mensa, Michaely and Schmalz (2014) for corroborating empirical evidence connecting payouts and agency).

The scenario in Column (2) raises average productivity, μ, 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, Φ, 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 2.30 to 2.49.

B.2. 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 Φ 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 parameter λ to 0.055 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 detailed analysis to future work.

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 newly derive a 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 model calibrated to pertinent US data finds that agency conflicts are much more severe than financial frictions.

Our novel 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.

REFERENCES

Abel, Andrew B., "The Effects of q and Cash Flow on Investment in the Presence of Measurement Error," *Journal of Financial Economics*, 2018, *128* (2), 363–377.

____ **and Janice C. Eberly**, "A Unified Model of Investment Under Uncertainty," *American Economic Review*, 1994, *84* (1), 1369–1384.

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," *Review of Economic Studies*, forthcoming.

Ai, Hengjie and Rui Li, "Investment and CEO Compensation under Limited Commitment," *Journal of Financial Economics*, 2015, *116* (3), 452–472.

Albuquerque, Rui and Hugo Hopenhayn, "Optimal Lending Contracts and Firm Dynamics," *Review of Economic Studies*, 2004, *71*, 285–315.

Altinkiliç, Oya and Robert S. Hansen, "Are There Economies of Scale in Underwriting Fees?," *Review of Financial Studies*, 2000, *13* (1), 191–218.

Asker, John, Allan Collard-Wexler, and Jan de Loecker, "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 2014, *122* (5), 1013–1063.

Bates, Thomas W., Kathleen M. Kahle, and René M. Stulz, "Why Do U.S. Firms Hold So Much More Cash Than They Used To?," *Journal of Finance*, 2009, *64* (5), 1985–2021.

Belo, Frederico, Xiaoji Lin, and Fan Yang, "External Equity Financing Shocks, Financial Flows, and Asset Prices," *Review of Financial Studies*, 2018. forthcoming.

Biais, Bruno, Thomas Mariotti, Guillaume Plantin, and Jean-Charles Rochet, "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications," *Review of Economic Studies*, 2007, *74*, 345–390.

Bils, Mark, Peter J. Klenow, and Cian Ruane, "Misallocation or Mismeasurement?," 2020. NBER Working Paper No. 26711.

Bolton, Patrick, Hui Chen, and Neng Wang, "A Unified Theory of Tobin's q, Corporate Investment, Financing, and Risk Management," *Journal of Finance*, 2011, *66* (5), 1545–1578.

____ **,** ____ **, and** ____ , "Market Timing, Investment, and Risk Management," *Journal of Financial Economics*, 2013, *109*, 40–62.

Boualam, Yasser, "Credit Markets and Relationship Capital," 2019. Working paper.

____ **, João F. Gomes, and Colin Ward**, "Understanding the Behavior of Distressed Stocks," 2020. Working paper.

Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin, "Finance and Development: A Tale of Two Sectors," *American Economic Review*, 2011, *101* (5), 1964–2002.

Calomiris, Charles W. and Margarita Tsoutsoura, "Underwriting Costs of Seasoned Equity Offerings: Cross-Sectional Determinants, Technological Change, and Pricing Benefits, 1980–2008," 2013. Working Paper.

Chen, Hui, "Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure," *The Journal of Finance*, 2010, *65* (6), 2171–2212.

Clementi, Gian Luca and Hugo A. Hopenhayn, "A Theory of Financing Constraints and Firm Dynamics," *Quarterly Journal of Economics*, 2006, *121*, 229–265.

DeMarzo, Peter M. and Michael Fishman, "Agency and Optimal Investment Dynamics," *Review of Financial Studies*, 2007, *20*, 151–188.

——— **and Yuliy Sannikov**, "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model," *Journal of Finance*, 2006, *61* (6), 2681–2724.

——— **, Michael J. Fishman, Zhiguo He, and Neng Wang**, "Dynamic Agency and the q Theory of Investment," *Journal of Finance*, 2012, *67* (6), 2295–2340.

Donaldson, Gordon, *Managing Corporate Wealth*, New York: Praeger, 1984.

Dow, James, Gary Gorton, and Arvind Krishnamurthy, "Equilibrium Investment and Asset Prices under Imperfect Corporate Control," *American Economic Review*, 2005, *95* (3), 659–681.

Eberly, Janice, Sergio Rebelo, and Nicolas Vincent, "Investment and Value: A Neoclassical Benchmark," 2009. NBER Working Paper No. 13866.

Edmans, Alex, Xavier Gabaix, and Augustin Landier, "A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium," *Review of Financial Studies*, 2008, *22* (12), 4881–4917.

Erickson, Timothy and Toni M. Whited, "Measurement Error and the Relationship between Investment and q," *Journal of Political Economy*, 2000, *108* (5), 1027–1057.

Fama, Eugene F. and Kenneth R. French, "Financing Decisions: Who Issues Stock?," *Journal of Financial Economics*, 2005, *76*, 549–582.

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

Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen, "Financing Constraints and Corporate Investment," *Brookings Papers On Economic Activity*, 1988, (1), 141–206.

Gillan, Stuart L., Jay C. Hartzell, Andrew Koch, and Laura T. Starks, "Getting the Incentives Right: Backfilling and Biases in Executive Compensation Data," *Review of Financial Studies*, 2018, *31* (4), 1460–1498.

Glover, Brent and Oliver Levine, "Idiosyncratic Risk and the Manager," *Journal of Financial Economics*, 2017, *126* (2), 320–341.

Gomes, Joao F., "Financing Investment," *American Economic Review*, 2001, *91* (5), 1263–1285.

Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez, "Capital Allocation and Productivity in South Europe," *Quarterly Journal of Economics*, 2017, *132* (4), 1915–1967.

Hall, Robert E., "Struggling to Understand the Stock Market," *American Economic Review*, 2001, *91* (2), 1–11.

Haltiwanger, John, Robert Kulick, and Chad Syverson, "Misallocation Measures: The Distortion that Ate the Residual," 2018. NBER Working Paper No. 24199.

Hartman-Glaser, Barney, Simon Mayer, and Konstantin Milbradt, "Corporate Liquidity Management under Moral Hazard," 2019. Working Paper.

Hayashi, Fumio, "Tobin's Marginal q and Average Q: A Neoclassical Formulation," *Econometrica*, 1982, *50*, 215–224.

Hennessey, Christopher A. and Toni M. Whited, "How Costly is External Financing? Evidence from a Structural Estimation," *Journal of Finance*, 2007, *62*, 1705–1745.

Hopenhayn, Hugo A., "Entry, Exit, and Firm Dynamics in Long-Run Equilibrium," *Econometrica*, 1992, *60* (5), 1127–1150.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Jensen, Michael C., "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers," *American Economic Review, Papers and Proceedings*, 1986, pp. 323–329.

____ **and William H. Meckling**, "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure," *Journal of Financial Economics*, 1976, pp. 305–360.

Kermani, Amir and Yueran Ma, "Asset Specificity of Non-Financial Firms," 2020. Working Paper.

King, Robert G. and Ross Levine, "Finance and Growth: Schumpeter Might Be Right," *Quarterly Journal of Economics*, 1993, *108* (3), 717–737.

Larcker, David F., "The Association Between Performance Plan Adoption and Corporate Capital Investment," *Journal of Accounting and Economics*, 1983, *5*, 3–30.

Lintner, John, "Distribution of Incomes of Corporations Among Dividends, Retained Earnings, and Taxes," *American Economic Review*, 1956, *46* (2), 97–113. Papers and Proceedings of the Sixty-Eighth Annual Meeting of the American Economic Association.

Lustig, Hanno, Chad Syverson, and Stijn Van Nieuwerburgh, "Technological Change and the Growing Inequality in Managerial Compensation," *Journal of Financial Economics*, 2011, *99*, 601–627.

McKeon, Stephen B., "Employee Option Exercise and Equity Issuance Motives," 2015. Working Paper.

Midrigan, Virgiliu and Daniel Yi Xu, "Finance and Misallocation: Evidence from Plant-Level Data," *American Economic Review*, 2014, *104* (2), 422–458.

Milbradt, Konstantin and Martin Oehmke, "Maturity Rationing and Collective Short-Termism," *Journal of Financial Economics*, 2015, *118* (3), 553–570.

Moll, Benjamin, "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?," *American Economic Review*, 2014, *104* (10), 3186–3221.

Morellec, Erwan, Boris Nikolov, and Norman Schürhoff, "Agency Conflicts Around the World," *Review of Financial Studies*, forthcoming.

Myers, Stewart C. and Nicholas S. Majluf, "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have," *Journal of Financial Economics*, 1984, *13*, 187–221.

Nikolov, Boris and Toni M. Whited, "Agency Conflicts and Cash: Estimates from a Dynamic Model," *Journal of Finance*, 2014, *69* (5), 1883–1921.

Quadrini, Vincenzo, "Investment and Liquidation in Renegotiation-Proof Contracts with Moral Hazard," *Journal of Monetary Economics*, 2004, *51*, 713–751.

Rajan, Raghuram G. and Luigi Zingales, "Financial Dependence and Growth," *American Economic Review*, 1998, *88* (3), 559–586.

Rampini, Adriano and S. Viswanathan, "Collateral, Risk Management, and the Distribution of Debt Capacity," *Journal of Finance*, 2010, *65*, 2293–2322.

Restuccia, Diego and Richard Rogerson, "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 2008, *11* (4), 707–720.

Riddick, Leigh A. and Toni M. Whited, "The Corporate Propensity to Save," *Journal of Finance*, 2009, *64* (4), 1729–1766.

Sannikov, Yuliy, "A Continuous-Time Version of the Principal-Agent Problem," *Review of Economic Studies*, 2008, *75* (3), 957–984.

Spear, Stephen E. and Sanjay Srivastava, "On Repeated Model Hazard with Discounting," *Review of Economic Studies*, 1987, *54* (4), 599–617.

Taylor, Lucian A., "CEO Wage Dynamics: Estimates from a Learning Model," *Journal of Financial Economics*, 2013, *108*, 79–98.

Tong, Jincheng and Chao Ying, "A Dynamic-Agency Based Asset Pricing Theory with Production," 2019. Working Paper.

Ward, Colin, "Agency in Intangibles," 2019. Working Paper.

Whited, Toni M. and Guojun Wu, "Financial Constraints Risk," *Review of Financial Studies*, 2006, *19* (2), 531–559.

Zhu, John Y., "Optimal Contracts with Shirking," *Review of Economic Studies*, 2013, *80* (2), 812–839.

Zwiebel, Jeffrey, "Dynamic Capital Structure under Managerial Entrenchment," *American Economic Review*, 1996, *86* (5), 1197–.

A. Technical Appendix

A. First-Best Solution

In the first-best economy there are neither agency ($\lambda = 0$) 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:[9]

$$q^{FB} = \max_i \frac{(1-\tau_Y)\mu - g(i)}{r + \delta - i}. \tag{A1}$$

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

$$g'(i^{FB}) = q^{FB} = \frac{(1-\tau_Y)\mu - g(i^{FB})}{r + \delta - i^{FB}}. \tag{A2}$$

Finally, because our managers are relatively impatient, it is best to pay them immediately, leaving $P^{FB}(K, W) = q^{FB}K - wK$ to investors.

B. Details of Solution Method

We solve the partial differential equation in (24) with a finite difference method that approximates the function $p(c, w)$ on a two-dimensional non-rectangular grid: $c \in \{c_i(w_j)\}_{i=1}^{I^j}$ and $w \in \{w_j(c_i)\}_{j=1}^{J^i}$, where we define $\overline{w}(c_i) = w_{J^i}(c_i)$ and $\overline{c}(w_j) = c_{I^j}(w_j)$. Each set of grid points along j, $w_j(c_i)$, depend on the value of c_i, because of the boundary curve $\{\overline{w}(c_i)\}_{i=1}^{I}$. 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 Δ_c and Δ_w. For the approximation of the derivatives at the boundaries, there are three different cases:

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

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

$$p(c_{I^j+1}, w_{J^i+1}) \approx p(c_{I^j+1}, w_{J^i}) - \Delta_w \approx p(c_{I^j}, w_{J^i}) - \Delta_w + \Delta_c.$$

We describe our computational algorithm below:

[9] We assume $\mu < g(r + \delta)$ and $q^{FB} > l_K$ to have a well-defined problem.

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

2. Calculate the investment policy function in (25)

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

4. We update the value function through an implicit method that solves the vector $p^{b+1} = (p_{1,1}^{b+1}, \ldots, p_{1,J^1}^{b+1}, p_{2,1}^{b+1}, \ldots, p_{2,J^2}^{b+1}, \ldots, p_{I^J,J^I}^{b+1})'$ 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

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

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

$$\mathbf{Q} = \begin{bmatrix}
q^{ss}_{1,1} & q^{su}_{1,1} & 0 & \cdots & 0 & q^{us}_{1,1} & q^{uu}_{1,1} & 0 & \cdots & 0 & \cdots & 0 \\
q^{sd}_{1,2} & q^{ss}_{1,2} & q^{su}_{1,2} & \ddots & \vdots & q^{ud}_{1,2} & q^{us}_{1,2} & q^{uu}_{1,2} & \ddots & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & q^{sd}_{1,J^1} & q^{ss}_{1,J^1} & 0 & \cdots & \cdots & q^{ud}_{1,J^1} & q^{us}_{1,J^1} & \ddots & \vdots \\
q^{ds}_{2,1} & q^{du}_{2,1} & 0 & \cdots & 0 & q^{ss}_{2,1} & q^{su}_{2,1} & 0 & \cdots & 0 & \ddots & \vdots \\
q^{dd}_{2,2} & q^{ds}_{2,2} & q^{du}_{2,2} & \ddots & \vdots & q^{sd}_{2,2} & q^{ss}_{2,2} & q^{su}_{2,2} & \ddots & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & q^{dd}_{2,J^2} & q^{ds}_{2,J^2} & 0 & \cdots & \cdots & q^{sd}_{2,J^2} & q^{ss}_{2,J^2} & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & q^{sd}_{I^J,J^I} & q^{ss}_{I^J,J^I}
\end{bmatrix}.$$

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

   - $q^{ss}_{i,j} = -\max(\mathbb{E}_t[dw], 0)/\Delta_w + \min(\mathbb{E}_t[dw], 0)/\Delta_w - \max(\mathbb{E}_t[dc], 0)/\Delta_c + \min(\mathbb{E}_t[dc], 0)/\Delta_c - \mathbb{E}_t[dw^2]/\Delta_w^2 - \mathbb{E}_t[dc^2]/\Delta_c^2$
   - $q^{su}_{i,j} = \max(\mathbb{E}_t[dw], 0)/\Delta_w + \mathbb{E}_t[dw^2]/(2\Delta_w^2)$

- $q^{sd}_{i,j} = -\min(\mathbb{E}_t[dw], 0)/\Delta_w + \mathbb{E}_t[dw^2]/(2\Delta_w^2)$
- $q^{us}_{i,j} = \max(\mathbb{E}_t[dc], 0)/\Delta_c + \mathbb{E}_t[dc^2]/(2\Delta_c^2)$
- $q^{ds}_{i,j} = -\min(\mathbb{E}_t[dc], 0)/\Delta_c + \mathbb{E}_t[dc^2]/(2\Delta_c^2)$
- $q^{uu}_{i,j} = -q^{du}_{ij} = -q^{ud}_{ij} = q^{dd}_{ij} = \mathbb{E}_t[dwdc]/(4\Delta_c\Delta_w)$,

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

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

$$B = \begin{bmatrix} (q^{dd}_{1,1} + q^{ud}_{1,1} + q^{sd}_{1,1}) \times (l_K + c_1) \\ \vdots \\ (q^{su}_{1,J^1}) \times (-\Delta_w) \\ (q^{dd}_{2,1} + q^{ud}_{2,1} + q^{sd}_{2,1}) \times (l_K + c_2) \\ \vdots \\ (q^{su}_{2,J^2}) \times (-\Delta_w) \\ \vdots \\ \vdots \\ (q^{su}_{I^J,J^I}) \times (-\Delta_w) \end{bmatrix} + \begin{bmatrix} (q^{ds}_{1,1} + q^{dd}_{1,1} + q^{du}_{1,1}) \times p_{1,1} \\ (q^{ds}_{1,2} + q^{dd}_{1,2} + q^{du}_{1,2}) \times p_{1,2} \\ \vdots \\ (q^{ds}_{1,J^1} + q^{dd}_{1,J^1} + q^{du}_{1,J^1}) \times p_{1,J^1} \\ \vdots \\ (q^{us}_{I^1,1}) \times (\Delta_c) \\ \vdots \\ \vdots \\ (q^{us}_{I^J,J^I}) \times (\Delta_c) \end{bmatrix},$$

and intuitively captures the rates of cash outflows from payments to managers, $-\Delta_w$, and inflows to investors from payouts, Δ_c, liquidation, $l_K + c$, and refinancing, $p_{1,j}$ for $j = 1, \ldots, J^1$, where we use the equilibrium value matching condition for refinancing, (11).

B.1. Stationary Distribution

The stationary distribution, $h(c,w)$, is calculated by solving, $h(c,w) = -(\mathbf{Q}^T)^{-1}\psi$, where ψ is the entry vector. The rows of ψ 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 $m = -\sum_i \mathbf{Q}^T h(c,w)\Delta_w\Delta_c$.

B.2. Accuracy of Solution

We depict the accuracy of these properties in Figure A-1. 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.

C. 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 (25) 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

$$g'(i(\overline{c}, \overline{w})) = p(\overline{c}, \overline{w}) + \overline{w} - \overline{c}, \text{ and} \tag{A3}$$

$$g'(i(\overline{c} + dc, \overline{w} + dw)) = p(\overline{c} + dc, \overline{w} + dw) + (\overline{w} + dw) - (\overline{c} + dc). \tag{A4}$$

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

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

which we can rearrange to yield

$$p(\overline{c}, \overline{w}) = p(\overline{c}, \overline{w} + dw) + dw = p(\overline{c} + dc, \overline{w} + dw) - dc + dw \tag{A5}$$

and then adding $\overline{w} - \overline{c}$ to both sides of (A5) gives

$$p(\overline{c}, \overline{w}) + \overline{w} - \overline{c} = p(\overline{c} + dc, \overline{w} + dw) + (\overline{w} + dw) - (\overline{c} + dc). \tag{A6}$$

Finally, the optimal investment rates along the boundary ((A3) and (A4)) and (A6) imply

$$i(\overline{c}, \overline{w}) = i(\overline{c} + dc, \overline{w} + dw).$$

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

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

Proposition (Tradeoff Along the Joint Upper Boundary)**.** *Consider a marginal change along the joint upper boundary from $(\overline{c}(w), \overline{w}(c))$ to $(\overline{c}(w) + dc, \overline{w}(c) + dw)$, 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 (24) is

$$
\begin{aligned}
rp(c,w) = \max_i p(c,w)(i-\delta) + p_c(c,w)\left((1-\tau_Y)\mu - g(i) + \delta\tau_Y + [r(1-\tau_C)-(i-\delta)]c\right) \\
+ p_w(c,w)(\gamma-(i-\delta))w + \frac{1}{2}p_{cc}(c,w)(\sigma(1-\tau_Y))^2 \\
+ \frac{1}{2}p_{ww}(c,w)\left(\frac{\sigma}{\mu}\lambda\right)^2 + p_{cw}(c,w)\frac{\sigma^2(1-\tau_Y)}{\mu}\lambda. \qquad \text{(A7)}
\end{aligned}
$$

Using our two points, $(\overline{c},\overline{w})$ and $(\overline{c}+dc,\overline{w}+dw)$, we can reduce (A7) to

$$
\begin{aligned}
p(\overline{c},\overline{w})[r-(i-\delta)] &= -(\gamma-(i-\delta))\overline{w} \\
&\quad + (1-\tau_Y)\mu - g(i) + \delta\tau_Y + [r(1-\tau_C)-(i-\delta)]\overline{c} \qquad \text{(A8)}
\end{aligned}
$$

and

$$
\begin{aligned}
p(\overline{c}+dc,\overline{w}+dw)[r-(i-\delta)] &= -(\gamma-(i-\delta))(\overline{w}+dw) \\
&\quad + (1-\tau_Y)\mu - g(i) + \delta\tau_Y + [r(1-\tau_C)-(i-\delta)](\overline{c}+dc), \qquad \text{(A9)}
\end{aligned}
$$

where we use the Lemma to simplify investment $i \equiv i(\overline{c},\overline{w}) = i(\overline{c}+dc,\overline{w}+dw)$. Therefore, when subtracting (A8) from (A9) we are left with

$$
[r-(i-\delta)](p(\overline{c}+dc,\overline{w}+dw) - p(\overline{c},\overline{w})) = [r(1-\tau_C)-(i-\delta)]dc - (\gamma-(i-\delta))dw. \qquad \text{(A10)}
$$

We can then place (A5) into (A10) to find

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

which reduces to (26) in the text. □

D. 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 (34). We ignore the dependence of $\mu_c(c,w)$ on state variables in what follows.

$$
\begin{aligned}
&\mathbb{E}[\Delta c\mathbf{1}\{\overline{c} < c+\Delta c < \mathcal{F}^{-1}(w)\}|c,w] \\
&= \int_{-\infty}^{\infty} \Delta c \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{1}{2}\left(\frac{\Delta c-\mu_c}{\sigma_c}\right)^2} \mathbf{1}\left\{\overline{c}(w) - c < \Delta c < \mathcal{F}^{-1}(w) - c\right\} d(\Delta c) \qquad \text{(A11)}
\end{aligned}
$$

We then use the change of variable $\Delta x = (\Delta c - \mu_c)/\sigma_c$ to change (A11) to

$$\int_{-\infty}^{\infty} (\sigma_c \Delta x + \mu_c) \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{1}{2}(\Delta x)^2} \mathbf{1}\left\{\overline{c}(w) - c < \sigma_c \Delta x + \mu_c < \mathcal{F}^{-1}(w) - c\right\} d(\sigma_c \Delta x + \mu_c)$$
$$= \int_{-\infty}^{\infty} (\sigma_c \Delta x + \mu_c) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\Delta x)^2} \mathbf{1}\left\{\frac{\overline{c}(w) - c - \mu_c}{\sigma_c} < \Delta x < \frac{\mathcal{F}^{-1}(w) - c - \mu_c}{\sigma_c}\right\} d(\Delta x)$$
$$= \int_{\frac{\overline{c}(w)-c-\mu_c}{\sigma_c}}^{\frac{\mathcal{F}^{-1}(w)-c-\mu_c}{\sigma_c}} (\sigma_c \Delta x + \mu_c) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\Delta x)^2} d(\Delta x)$$
$$= \frac{\sigma_c}{\sqrt{2\pi}} \int_{\frac{\overline{c}(w)-c-\mu_c}{\sigma_c}}^{\frac{\mathcal{F}^{-1}(w)-c-\mu_c}{\sigma_c}} \Delta x e^{-\frac{1}{2}(\Delta x)^2} d(\Delta x) + \mu_c \int_{\frac{\overline{c}(w)-c-\mu_c}{\sigma_c}}^{\frac{\mathcal{F}^{-1}(w)-c-\mu_c}{\sigma_c}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\Delta x)^2} d(\Delta x). \quad \text{(A12)}$$

Since under the change $u = -\frac{1}{2}x^2$ we have $\int x e^{-\frac{1}{2}x^2} dx = -\int e^u du = -e^u + k = -e^{-\frac{1}{2}x^2} + k$, where k is a constant of integration, equation (A12) equals

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

E. VERIFICATION OF HJB OPTIMALITY AND FULL EFFORT CONDITION

Define the gain process $\{\mathcal{G}\}$ 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

$$e^{rt} d\mathcal{G}_t = K_t \left\{ \begin{array}{c} \left[\begin{array}{c} -rp + p(i_t - \delta) + p_c\left((1-\tau_Y)\mu - g(i_t) + \delta\tau_Y + [r(1-\tau_C) - (i_t - \delta)]c_t\right) \\ +p_w(\gamma - (i_t - \delta))w_t + \frac{1}{2}p_{cc}(\sigma(1-\tau_Y))^2 + \frac{1}{2}p_{ww}\left(\beta_t(1-\tau_Y)\sigma\right)^2 + p_{cw}\beta_t(1-\tau_Y)^2\sigma^2 \end{array} \right] dt \\ + \left[(1-p_c) \times (dD_t - dF_t) - dX_t - (1+p_w)dU_t\right]/K_t + (p_c + \beta_t p_w)\sigma(1-\tau_Y)dZ_t \end{array} \right\},$$

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 $\beta_t^* = \lambda/\left((1-\tau_Y)\mu\right)$ the top two lines in the square brackets are the optimized PDE in (24) 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_{ww} and the sum $\beta p_{ww}/2 + p_{cw}$, under the policy with the smallest β, are nonpositive.[10] Panels A and B of Figure A-2 depict p_{ww} and the sum under β_t^* and the

[10]This can be seen by simplifying the second-order terms that depend on the incentive coefficient β,

calibration in Table I and show that these terms are negative across the entire state space.

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

$$d\mathcal{G}_t = \mu_{\mathcal{G}}(t)dt + e^{-rt}K_t(p_c + \beta_t p_w)\sigma(1-\tau_Y)dZ_t,$$

where $\mu_{\mathcal{G}}(t) \leq 0$. Let $\varphi_t \equiv e^{-rt}K_t(p_c+\beta_t p_w)\sigma(1-\tau_Y)$. 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 $\{\mathcal{G}\}$ is a supermartingale.

We can now evaluate the principal's payoff for an arbitrary incentive compatible contract. Recall that $P(K_\tau, C_\tau, W_\tau) = l_K K_\tau + l_C C_\tau$. Given any $t < \infty$,

$$\mathbb{E}\left[\int_0^\tau e^{-rs}(dD_s - dF_s - dX_s - dU_s) + e^{-r\tau}\left(l_K K_\tau + l_C C_\tau\right)\right]$$
$$= \mathbb{E}\left[\mathcal{G}_{t\wedge\tau} + 1_{t\leqslant\tau}\left(\int_t^\tau e^{-rs}(dD_s - dF_s - dX_s - dU_s) + e^{-r\tau}\left(l_K K_\tau + l_C C_\tau\right) - e^{-rt}P(K_t, C_t, W_t)\right)\right]$$
$$= \mathbb{E}\left[\mathcal{G}_{t\wedge\tau}\right] +$$
$$e^{-rt}\mathbb{E}\left[1_{t\leqslant\tau}\left(\mathbb{E}_t\left[\int_t^\tau e^{-r(s-t)}(dD_s - dF_s - dX_s - dU_s) + e^{-r(\tau-t)}\left(l_K K_\tau + l_C C_\tau\right)\right] - P(K_t, C_t, W_t)\right)\right]$$
$$\leq \mathcal{G}_0 + \left(q^{FB} - (l_K + c_t)\right)\mathbb{E}[e^{-rt}K_t].$$

The first term of the inequality follows from the nonpositive drift of $d\mathcal{G}_t$ and the martingale property of $\int_0^{t\wedge\tau} \varphi_s dZ_s$. The second term follows from

$$\mathbb{E}_t\left[\int_t^\tau e^{-r(s-t)}(dD_s - dF_s - dX_s - dU_s) + e^{-r(\tau-t)}\left(l_K K_\tau + l_C C_\tau\right)\right] \leq q^{FB}K_t - w_t K_t,$$

which is the first-best result and

$$q^{FB}K_t - w_t K_t - P(K_t, C_t, W_t) \leq (q^{FB} - p(c_t, 0))K_t = \left(q^{FB} - (l_K + c_t)\right)K_t.$$

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 $t \to \infty$

$$\mathbb{E}\left[\int_0^\tau e^{-rs}(dD_s - dF_s - dX_s - dU_s) + e^{-r\tau}\left(l_K K_\tau + l_C C_\tau\right)\right] \leq \mathcal{G}_0. \qquad \text{(A13)}$$

$p_{ww}\beta^2(1-\tau_Y)^2\sigma^2/2 + p_{cw}\beta(1-\tau_Y)^2\sigma^2$, to be less than or equal to zero.

for all incentive-compatible contracts. On the other hand, under the optimal contract $\mathcal{C}^*$, principal's payoff $\mathcal{G}(\mathcal{C}^*)$ achieves $\mathcal{G}_0$ because the above weak inequality holds with equality when $t \to \infty$.

E.1. Full Effort Condition

Finally, we require λ 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 λdt. 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

$$dW_t = \gamma W_t dt - dU_t - \lambda K_t dt.$$

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:

$$\begin{aligned} rp \geq\ & p(i - \delta) + p_c\left(-g(i) + \delta\tau_Y + [r(1 - \tau_C) - (i - \delta)]c\right) \\ & + p_w(\gamma - \lambda - (i - \delta))w + \frac{1}{2}p_{cc}(\sigma(1 - \tau_Y))^2, \text{ for all } c \text{ and } w. \end{aligned}$$

We confirm this optimality of full effort numerically and depict the result in Panels C and D of Figure A-2. 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.

F. 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

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

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

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

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

1. The termination boundary is

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

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

$$p(c,w) = p(c-(w-\overline{w}(c)), \overline{w}(c)) - (w-\overline{w}(c)) \text{ for } w \geq \overline{w}(c)$$
$$\Rightarrow \frac{p(c,w)-p(c,\overline{w}(c))}{w-\overline{w}(c)} = \frac{p(c-(w-\overline{w}(c)),\overline{w}(c)) - p(c,\overline{w}(c))}{w-\overline{w}(c)} - 1$$
$$\Rightarrow p_w(c,\overline{w}(c)) = -p_c(c,\overline{w}(c)) - 1, \text{ as } w \to \overline{w}(c), \quad \text{(A14)}$$

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 fK. 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

   $$p(0,w) = p(f,w) - \Phi - (1+\phi)f, \quad \text{(A15)}$$

   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

   $$p(0,w) = p(f,\overline{w}(f)) - \Phi - (1+\phi)f - (w-\overline{w}(f)). \quad \text{(A16)}$$

   Differentiating this equation with respect to c gives

   $$p_c(f,\overline{w}(f)) + p_w(f,\overline{w}(f))\frac{\partial \overline{w}(c)}{\partial c}\bigg|_{c=f} + \frac{\partial \overline{w}(c)}{\partial c}\bigg|_{c=f} = 1+\phi. \quad \text{(A17)}$$

   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 (A14) 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 (A17) we get

$$p_c(f, \overline{w}(f)) \left(1 - \frac{\partial \overline{w}(c)}{\partial c}\bigg|_{c=f}\right) = 1 + \phi. \tag{A18}$$

4. When cash holdings get large we have as before

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

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

$$\begin{aligned} p_{cw}(\overline{c}(w), w) &= 0 \text{ for each } w, \\ p_{cw}(c, \overline{w}(c)) &= -p_{cc}(c, \overline{w}(c)) \text{ for each } c\text{, and} \\ p_{cw}(\overline{c}(w), \overline{w}(c)) &= -p_{cc}(\overline{c}(w), \overline{w}(c)) = 0 \text{ for every } c \text{ and } w. \end{aligned}$$

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 $\overline{c}(w)$ the inequality collapses to $p_w(\overline{c}(w), w) \geq -2$: 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 ($p_c(\overline{c}(w), w) = 1$) and raising management's continuation utility costs, at most, another dollar. Because (A14) 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-3. The refinancing decision $f(w)$ is traced out with the dashed line. Refinancing decisions satisfy (A15) 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 (A17). The firm would decide by choosing the maximum of (A15) and (A16).

Of course, this is not the only possibility for refinancing. A different equilibrium could be imposed by simply requiring (A15) 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 summarize, 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 (A16) 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 further complexity into an already challenging setup to solve 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, 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 $p_w(\overline{c}(w), \overline{w}(c)) = -2$. Using this condition rather than $p_w(\overline{c}(w), \overline{w}(c)) = -1$ 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 $\max$(pstkr(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))

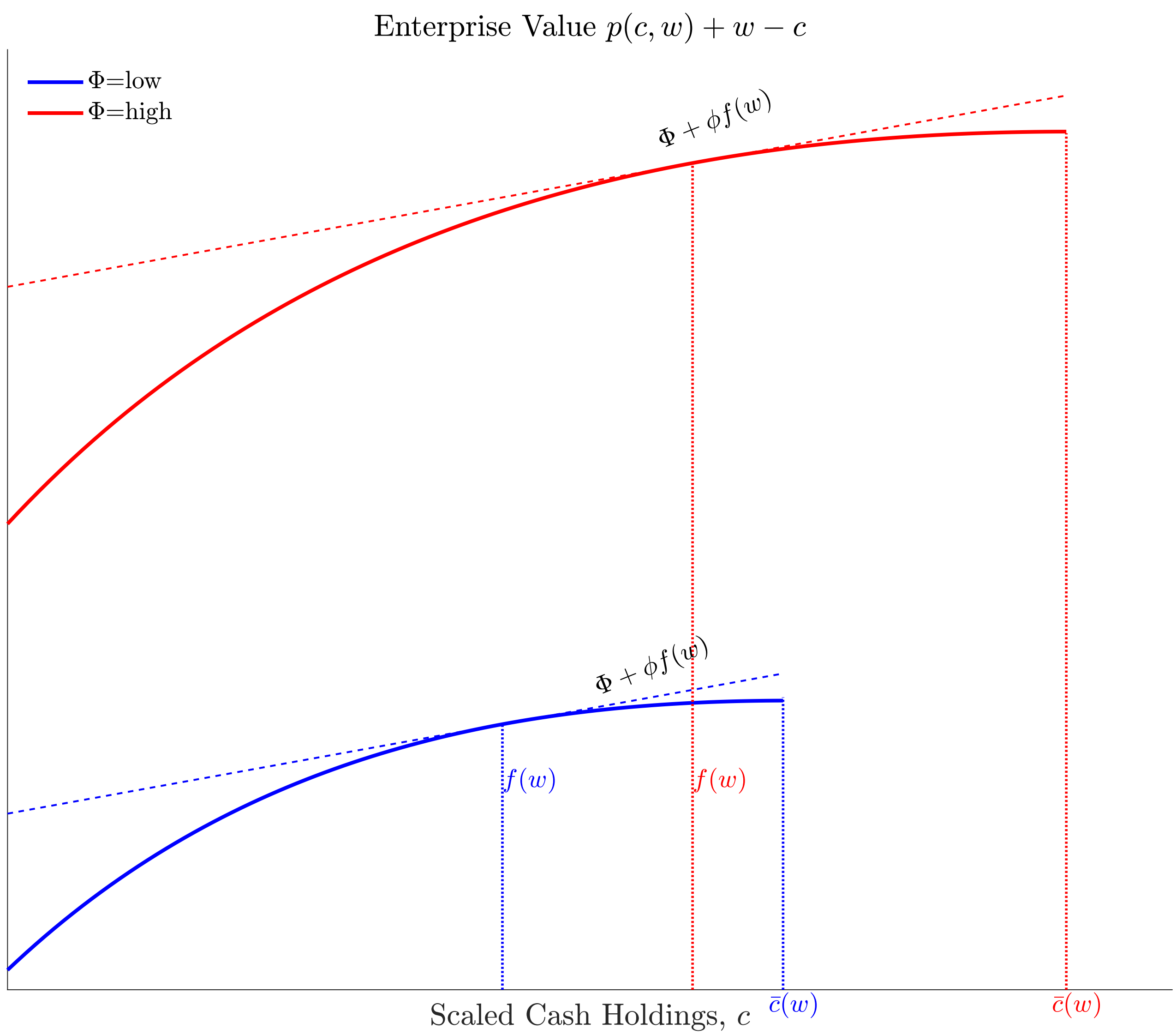

FIGURE 1: COST AND BENEFIT OF REFINANCING AND TOTAL ENTERPRISE VALUE

This figure depicts the refinancing decision under two parameterizations of Φ. The optimal decision equates the gain in enterprise value $(p(f(w),w)+w-f(w)-(p(0,w)+w)$ to the total refinancing cost $\Phi+\phi f(w)$ where $f(w)$ is determined by the condition $p_c(f(w),w)=1+\phi$. The payout boundaries are denoted by $\bar{c}(w)$.

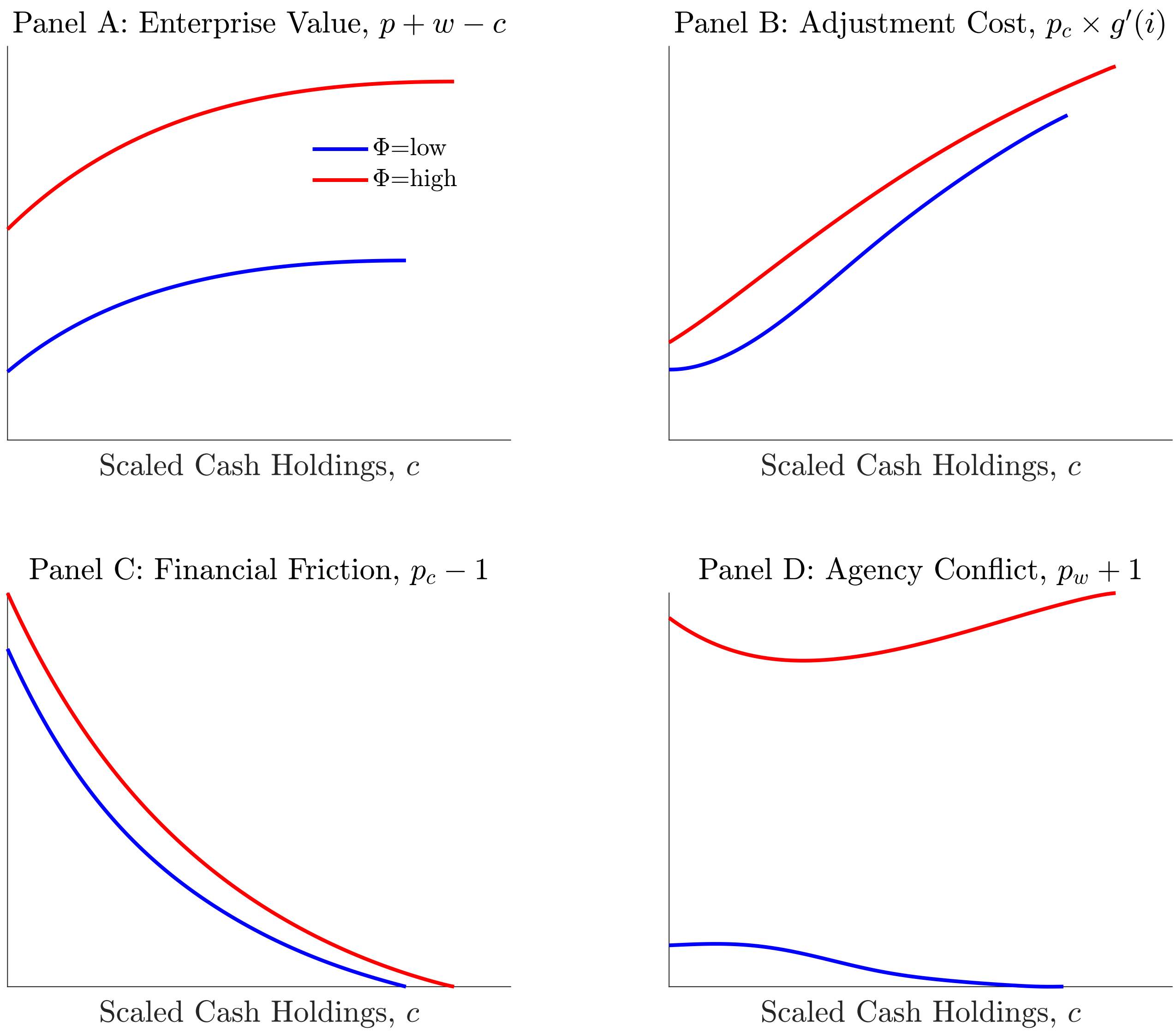

FIGURE 2: MARGINAL ENTERPRISE VALUE AND INVESTMENT

This figure decomposes the investment decision $p + w - c = p_c \times g'(i) + (p_c - 1)c + (p_w - 1)w$ under two values of Φ. Marginal enterprise value with respect to cash holdings is $\partial(p(c, w) + w - c)/\partial c = p_c - 1$ and with respect to managers" stake is $\partial(p(c, w) + w - c)/\partial w = p_w + 1$.

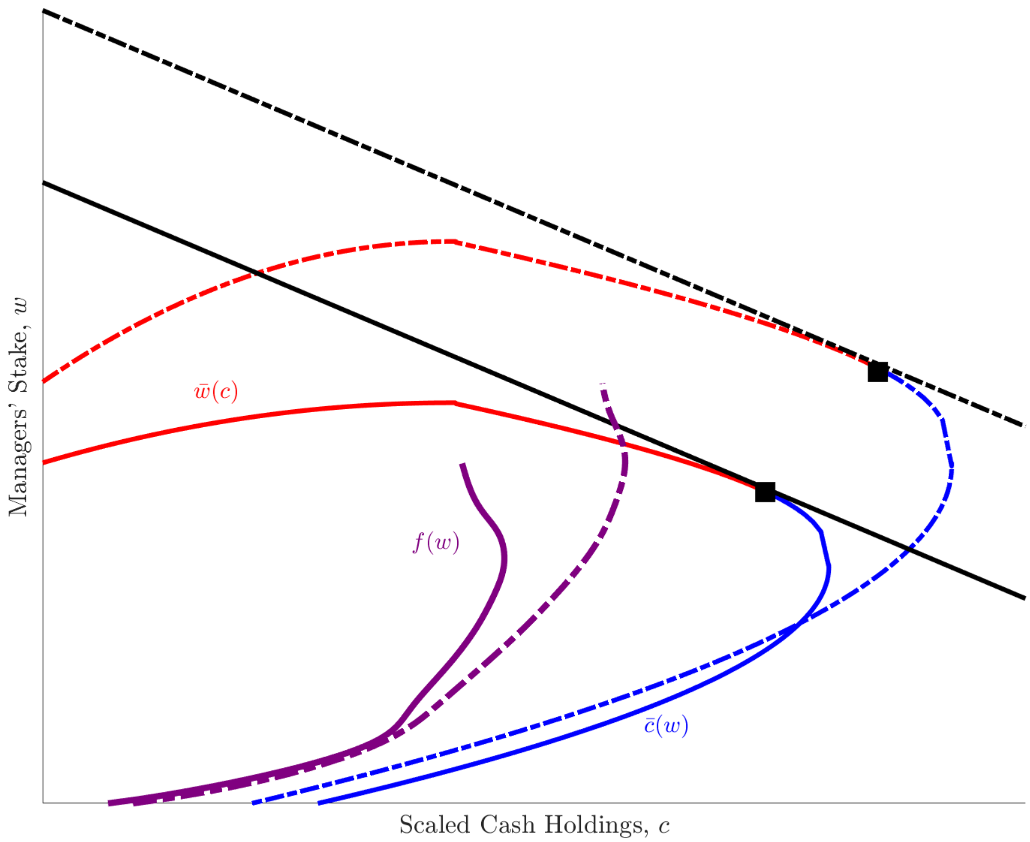

FIGURE 3: BOUNDARIES AND REFINANCING

This figure shows investors" payout boundary $\overline{c}(w)$, managers" payment boundary $\overline{w}(c)$, and the refinancing decision $f(w)$ for a low (solid lines) and high value (dashed lines) of Φ. Additionally, the joint upper boundaries $(c, w) = (\overline{c}(w), \overline{w}(c))$ is marked by a black square. Tangent to the joint upper boundary is the second-best frontier in black that has the slope of $-\tau_c r/(\gamma - r)$.

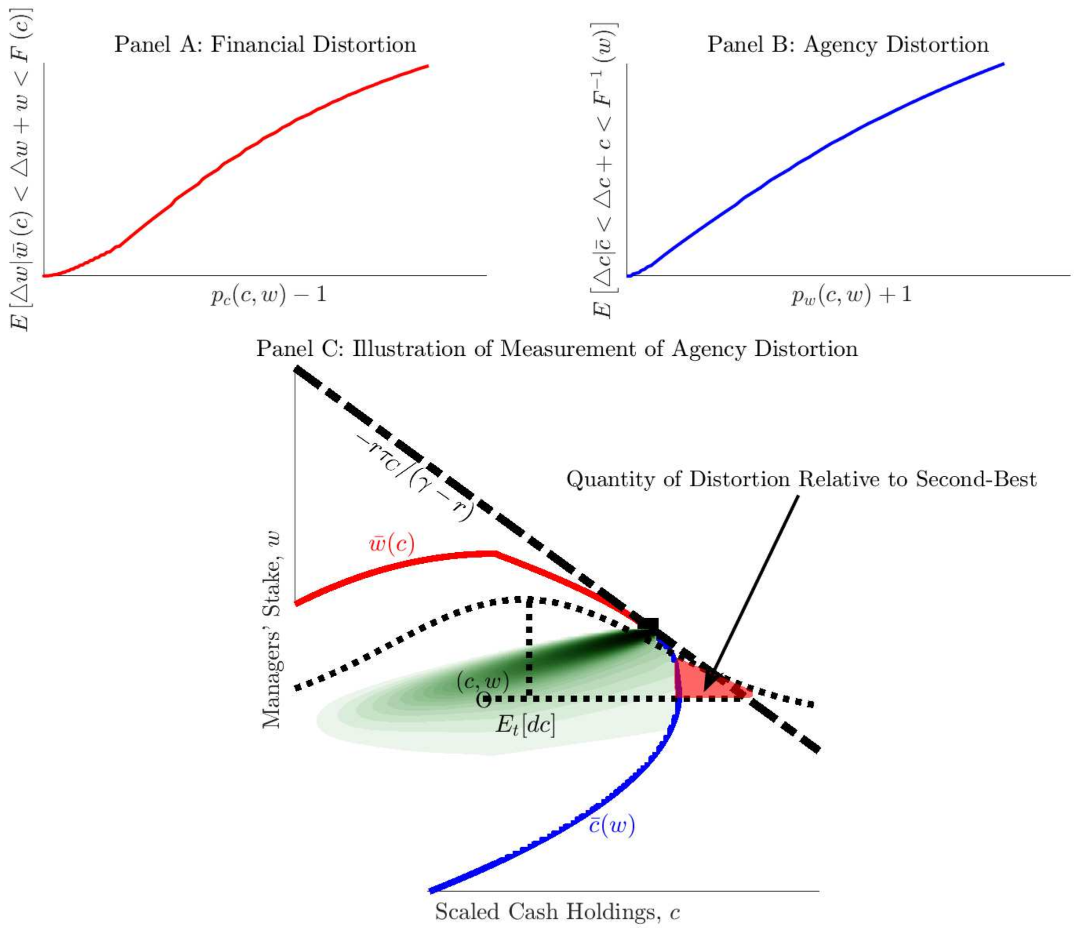

FIGURE 4: ILLUSTRATION OF MEASUREMENT OF DISTORTIONS

Panels A and B plot the financial and agency distortions on marginal enterprise value. Panel C 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, $\overline{w}(c)$, and the blue line the payout boundary, $\overline{c}(w)$. These boundaries intersect at the joint upper boundary, $(\overline{c}(w), \overline{w}(c))$, marked by the black square. The second-best frontier, the dash-dotted line, touches the joint upper boundary and has slope $dw/dc = -r\tau_C/(\gamma - r)$. 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 $\mathbb{E}_t[dc]$ 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 $\overline{c}(w)$ and ending at the frontier $\mathcal{F}^{-1}(w)$.

TABLE I: VARIABLE DEFINITIONS AND CALIBRATION

This table list the values of the illustrative calibration in Section III and the benchmark calibration discussed in Section IV. All parameters are annualized.

Parameter	Symbol	Value	
		Illustration	Calibration
Interest Rate	r	0.04	0.04
Average Productivity	μ	0.17	0.17
Volatility of Productivity	σ	0.40	0.35
Depreciation of Capital	δ	0.08	0.08
Adjustment Cost/Capital's Doubling Time in Years	θ	1	6
Exogenous Growth of Capital	z	0	0.09
Liquidiation Value of Capital	l_K	0.63	0.4
Management's Discount Rate	γ	0.048	0.048
Agency Costs	λ	0.04	0.06
Refinancing Cost (Fixed)	Φ	(0.008, 0.016)	0.005
Refinancing Cost (Variable)	ϕ	0.05	0.05
Corporate Tax Rate	τ_Y	0.3	0.3
Penalty on Cash Holdings	τ_C	0.2	0.2
Managers' Relative Bargaining Power	ψ	0.5	0.5
Average Entrant Cash Holdings	$\mathbb{E}[c_0]$	0.15	0.15

TABLE II: INTERNAL CALIBRATION TARGETS

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 $\mathbb{E}_t[dY]/K + \tau_Y \delta$. Indicator variables are payout rate (37), entry/exit rate (28), refinancing rate (36), and refinancing size which is $f(w)$.

	Model	Data			
	Mean	Mean	P25	P50	P75
Cash Holdings	21.6	22.7	2.1	7.8	26.2
Compensation	1.4	1.3	0.3	0.7	1.6
Investment	8.2	8.7	2.8	5.5	10.9
Free Cash Flow	3.9	4.8	-0.2	7.8	14.6
Payout Rate	37.6	53.2			
Entry/Exit Rate	1.7	2.0			
Refinancing Rate	16.4	16.8			
Refinancing Size	13.9	51.5	12.4	24.8	58.9

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 $\mathbb{E}_t[dY]/K + \tau_Y \delta$. Indicator variables are payout rate (37), entry/exit rate (28), refinancing rate (36), and refinancing size which is $f(w)$. Agency and financial distortions are computed respectively in (33) and (34).

	Benchmark	$\mu = 0.18$	$\tau_Y = 0.21$	$\Phi = 0.01$	$\tau_Y = 0.21$, $\Phi = 0.01$	$\lambda = 0.055$, $\psi = 0.03$
	(1)	(2)	(3)	(4)	(5)	(6)
Cash Holdings	21.6	25.6	22.5	24.5	25.5	24.7
Compensation	1.4	1.5	1.6	1.5	1.7	1.6
Investment	8.2	9.2	8.9	8.3	9.2	8.9
Free Cash Flow	3.9	4.2	4.5	3.9	4.3	3.6
Payout Rate	37.6	35.2	39.3	35.2	36.9	36.2
Entry/Exit Rate	1.7	1.8	0.2	2.4	0.4	0.4
Refinancing Rate	16.4	13.2	18.2	14.1	14.9	14.1
Refinancing Size	13.9	17.0	16.0	17.3	17.1	18.9
Distortions						
Agency (a)	2.30	2.51	1.56	2.49	1.81	1.62
Financial (f)	0.46	0.42	0.59	0.44	0.56	0.45
Ratio (a/f)	5.05	6.03	2.63	5.60	3.25	3.56
Relative to (1)	1.00	1.19	0.52	1.11	0.64	0.71

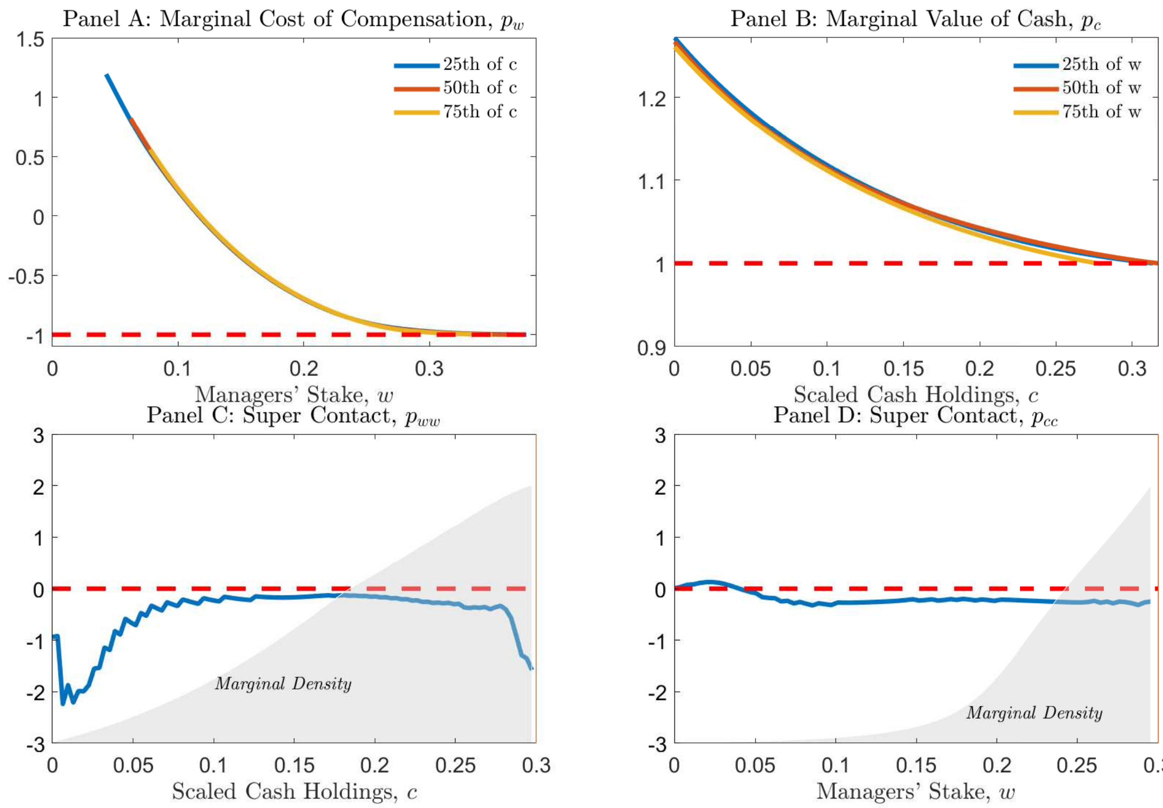

FIGURE A-1: ACCURACY OF MODEL SOLUTION

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.

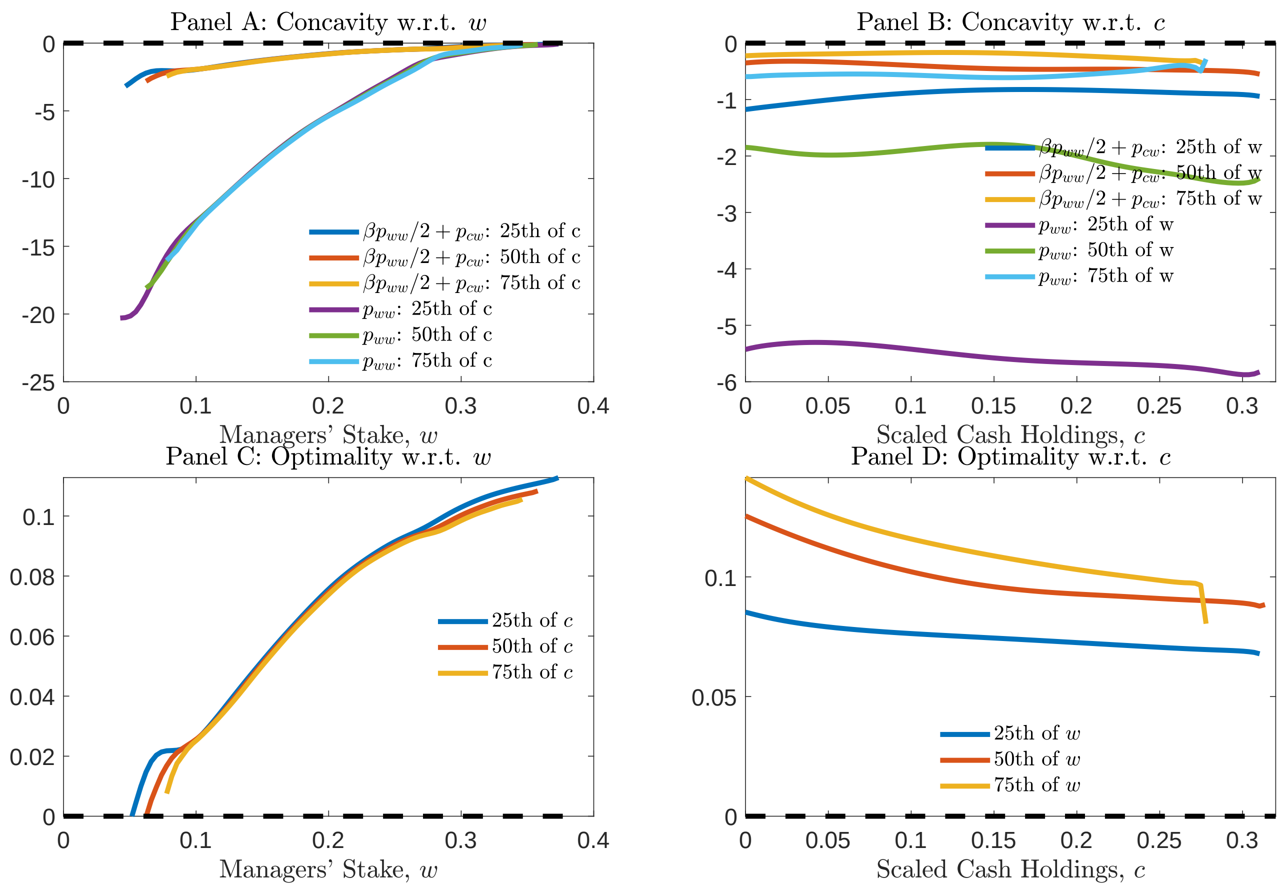

FIGURE A-2: CONCAVITY AND OPTIMALITY OF THE HJB EQUATION

This figure displays in Panels A and B the concavity of the value function required for $\beta \geq \lambda/(\mu(1-\tau_Y))$ 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 to 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.

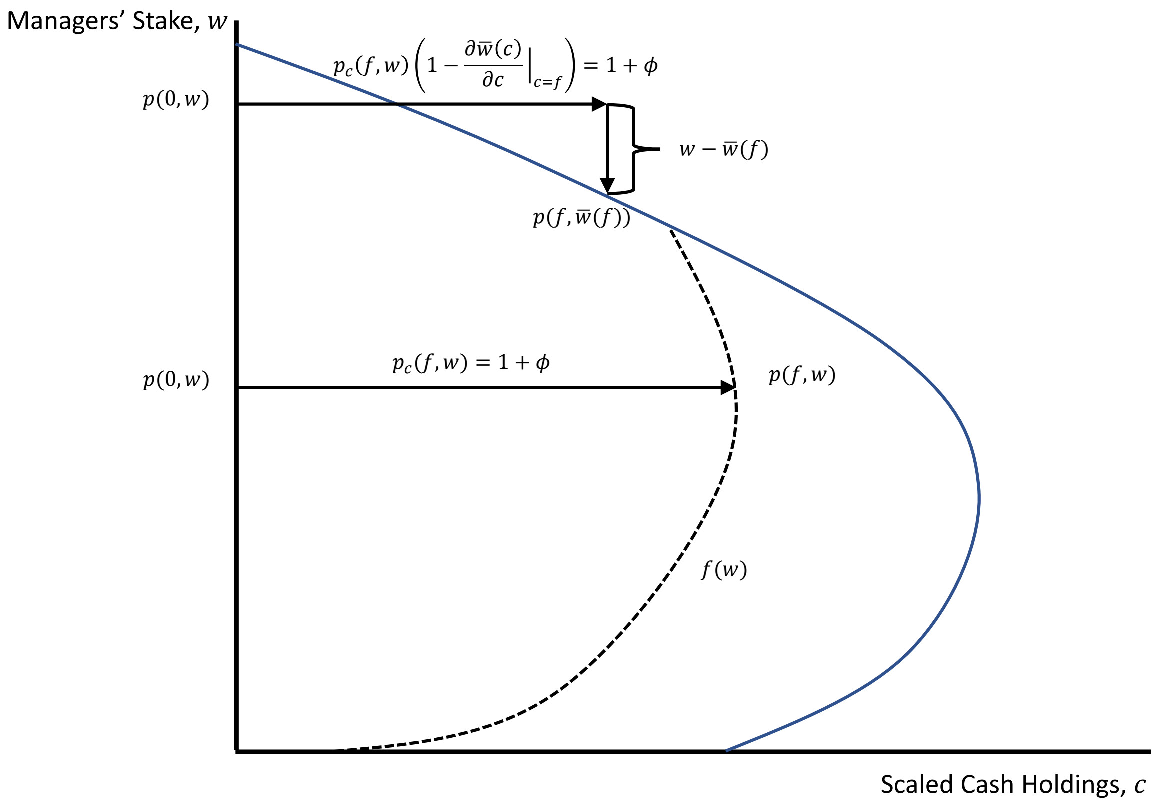

FIGURE A-3: HYPOTHETICAL ILLUSTRATION OF ALTERNATIVE SETUP

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.