Firm Investment and Stakeholder Choices: A Top-Down Theory of Capital Budgeting*

Andres Almazan
University of Texas

Zhaohui Chen
University of Virginia

Sheridan Titman
University of Texas and NBER

28 May 2011

Abstract

This paper provides a model of the firm's capital budgeting process that considers the interaction of top executives that have private information about the firm's prospects; and lower level managers (and other stakeholders) whose motivation and efforts affects the likelihood of the firm's success. In this setting, top executives' decisions affect lower level managers inferences about firm's prospects. Specifically, higher levels of investment commitments indicate that the firm has promising prospects and induce stakeholders to take actions that contribute to the firm's success. In this setting, we examine the role of commonly observed capital budgeting rules which not only play their traditional allocative role but also affect how private information is transmitted from the top down. As we show, in this framework, there can arise a number of commonly observed investment distortions such as capital rationing, investment rigidities, overinvestment, and inflated discount rates.

*We would like to thank Neal Galpin, Adolfo de Motta and David Dicks and seminar participants at ANU, DePaul University, the 2011 Finance Down Under Conference, UCLA, Universidad Carlos III, Chinese University of Hong Kong, University of Essex, University of Melbourne, Nanyang Technological University, National University of Singapore, UNC-Chapel Hill, University of Notre Dame, University of Pennsylvania, University of Virginia and UT-Austin. Corresponding author: andres.almazan@mccombs.utexas.edu.

1 Introduction

The capital budgeting process has been described (i.e., Brealey, Myers and Allen 2008) as a combination of "bottom-up" procedures, where lower units solicit capital from headquarters, and "top-down" procedures where headquarters use their discretion to allocate capital downstream. An extensive literature has analyzed the incentive and information considerations that can emerge in "bottom-up" capital allocation processes (e.g., Harris and Raviv 1996 or Bernardo et al. 2004), however, up to now, the literature has not focused on "top-down" processes in which capital allocation decisions are made by better informed headquarters.

In reality top-down procedures are likely to be relevant even in settings in which headquarters receive requests for funds from downstream managers. By aggregating the information contained in such requests, headquarters end up acquiring broader information which helps the firm determine its overall investment expenditures, (e.g., the information provided by one unit has implications for another unit's investment). As we discuss in this paper, in these situations, the capital allocation choices of headquarters can convey information to the managers of the lower units as well as to the other stakeholders of the firms. When this is the case, the procedures that firms use to evaluate investments can influence how this information is transmitted and how lower units interpret such information.[1]

As an illustration of the importance of the information conveyed by investment choices, consider the capital allocation decision of a large integrated oil company, such as Royal Dutch Shell. For a firm like this, a major investment in biofuels is likely to be viewed as an indication that the firm's top management has favorable information about the future prospects of these alternative sources of energy or, alternatively, that it has serious concerns about the future of the more traditional sources of energy.[2] In response to the favorable information conveyed by the investment choice, employees in Shell's biofuel division may be encouraged to work harder which, in turn, make the biofuel division more successful.

[1]Supporting the importance of top-down information transmission, Grinstein and Tolowsky (2004) provide evidence that suggests that directors of S&P 500 firms act to "alleviate conflicts of interest between agents and principals and to communicate principals' information to agents."

[2]In February 2010, Shell signed an agreement to create a joint venture to distribute and sell cane ethanol in Brazil and committed to contribute about \$1.6 billion in cash and other assets. Subsequently analysts stated that "*Shell has now made a major commitment to cane ethanol in particular and biofuels in general.*" (See http://247wallst.com/2010/08/26/shell-cosan-jv-highlights-ethanol-rds-a-czz-cdxs-adm-vlo-peix/).

To better understand how capital budgeting procedures and stakeholder perceptions interact we develop a simple model of a firm whose production process requires a capital expenditure (i.e., an investment) as well as effort exerted by a stakeholder (i.e., an employee). In our setting, the firm's owner (i.e., the entrepreneur) first obtains private information about the firm's prospects, and then chooses the level of investment and the employee's (pay-for-performance) compensation. The employee's effort is more productive when the future prospects of the firm are better which, in combination with his compensation contract, induces the employee to exert more effort when his beliefs about the firm's future prospects are more favorable. As a result, the entrepreneur has the incentive to overinvest (relative to the case of symmetric information) because doing so conveys favorable information, which in turn, elicits additional effort from the employee.

Although we also explore a model with a continuum of firm types, most of our model can be illustrated in a setting in which the firm's prospects are one of two types, high or low. Within this setting two alternative firm policies emerge as optimal: (i) a "separation policy" in which a high prospect firm invests more (and compensates the manager more) than a low prospect firm and (ii) a "pooling policy" in which regardless of their types, firms commit to a fixed level of investment and compensation. With the separation policy, the firm invests more when the marginal productivity of capital is higher but tends to overinvest because of the potential benefits associated with conveying favorable information. With the pooling policy, investment is independent of the firm's information but the firm's overinvestment incentive is suppressed. In other words, the choice between the pooling and the separation policy is determined by a trade-off between the efficiency gain associated with having an investment policy which incorporates information and the efficiency loss associated with overinvestment.

Having established the incentive to overinvest and the potential benefits of committing to a fixed level of investment we enrich the previous framework by adding a third party to our model. In addition to the entrepreneur, who obtains information, and the employee who exerts effort, we consider a third party that sets capital budgeting policies and offers a compensation contract to the entrepreneur. One interpretation is that the third party is a venture capitalist that provides funding for the firm and retains some control of its operations. Another is that the informed party is the manager of the division of a conglomerate and the third party is either the executives at the firm's headquarters or the firm's board.

Within this setting, we investigate a number of issues related to the capital budgeting process. First, we consider how the firm's capital structure can be designed to reduce incentives to overinvest. Second, we examine the possibility of commiting to using a specific hurdle rate that may not equal the rate that would be used with symmetric information. (This could be a single discount rate that is independent of the investment expenditure or it could be a function of investment expenditures.) Third, we examine whether the possibility of offering a compensation contract to the entrepreneur affects the firm's use of the capital budgeting policies to convey information. Specifically, we consider EVA-like compensation contracts whereby compensation is determined by the firm's cashflow minus a pre-specified capital cost. Finally, we examine under which conditions suppressing (or reducing) pay-for-performance compensation to the entrepreneur can increase firm value.

The analysis shows that several commonly observed capital budgeting practices can enhance firm value. First, the use of debt financing can create a Myers (1977) debt overhang problem that offsets the incentive to overinvest. In addition, the adoption of high hurdle rates can change the opportunity cost of capital for the entrepreneur and hence offset the incentive to overinvest that arises when the level of investment conveys information to stakeholders.[3] Finally, compensation contracts can directly change the entrepreneur's incentives to use the firm investment policy to communicate information to the stakeholders.

As we mentioned at the outset, our analysis of a "top-down" capital allocation process is in contrast of the analysis of "bottom-up" process that is the focus of the existing literature. Specifically, the capital budgeting literature has examined the issue of how a firm may distort its capital budgeting practices in order to induce managers to exert proper effort (Bernardo, Cai, and Luo 2001, 2004, 2006), to curb managers' empire building tendencies (e.g., Harris and Raviv 1996, 1997, Marino and Matsusaka 2004, and Berkovitch and Israel 2004) or to reveal their private information. This literature has also considered the trade-offs that arise in the decision to delegate capital budgeting decisions to a better informed agent (e.g., Aghion and Tirole, 1997 and Burkart, Gromb, and Panunzi, 1997).

Our analysis also has implications that are similar to the agency literature that argues

[3]See Poterba and Summers (1995) who document the use of high-discount rates on American corporations and Meier and Tarhan (2007) who provide evidence of what they refer to as a "hurdle rate puzzle" i.e., the use of a discount rate that substantially exceeds estimates of their cost of capital.

that since managers get private benefits from managing larger enterprises, shareholders may want to impose restrictions on managers that inhibit their incentives to overinvest (e.g., Jensen 1986, and Hart and Moore 1995). As we show, the tendency to overinvest, as well as procedures that curb this tendency, can also arise within a setting without managerial private benefits. Hence, in addition to offering a theory that can rationalize investment rigidities in large corporations, we offer an explanation for the high hurdle rates imposed by venture capitalists on the investment choices of young start-up firms.[4]

Although the setting is very different, our analysis is actually closest to Hermalin (1998) and Komai, Stegeman, and Hermalin (2007), who study the problem of leadership in organizations.[5] In these models, an informed leader signals his favorable information by expending greater effort, which in turn motivates his subordinates to work harder (i.e., leadership effects). There are, however, a number of key differences between our approach and the setting considered in the leadership papers. First, the entrepreneur's costly action, in our setting the investment choice, can be contractible, while the leader's effort in the leadership models is not. Indeed, the capital budgeting policies that we explore arise because of the contractibility of investment. Second, these papers focus on how to ameliorate undereffort in teams while our model concentrates on how to design capital investment rules that effectively take advantage (and sometimes limit) leadership effects. Third, since we consider an unique stakeholder, we abstract from any team related effects and consider explicit performance contracts to influence the choice of effort by agents. Finally, in the last part of our analysis, we introduce a third party that is not involved in the production process but which has the authority to set investment rules and to claim firm output. As we show, the presence of a third party who acts as "budget breaker" can broaden the contract space, and thus provide the entrepreneur with incentives that lead to more efficient investment choices.[6]

[4]Indeed, private equity firms and venture capitalist take this tendency to evaluate investments with very high discount rates to an extreme, generally requiring "expected" internal rates of return on their new investments that exceed 25%. Gompers (1999) describes the "venture capital" valuation method, where discount rates of more than 50% per year are used.

[5]See also Benabou and Tirole (2003) who consider the signaling effect of providing explicit incentives to employees.

[6]Holmstrom (1982) first pointed out the value of third parties who by threatening to break the budget eliminate inefficiencies in team production. While in Holmstrom (1982), third parties act "off-equilibrium path" in our setting, the third party breaks the budget in equilibrium in order to improve entrepreneurial incentives to invest appropriately.

More generally, our paper belongs to the principal-agent literature with informed principals. Among other things, this literature considers the effects of the principal's information on the optimal compensation contract (Beaudry 1994 and Inderst 2000), the value of the private information to the principal (Chade and Silvers 2004 and Karle 2009) and the incentives to disclose information (i.e., provide "advice") to the agent (Strausz 2009). In contrast to these models, the principal in our model not only designs the agent's compensation but also takes an action (i.e., makes an investment) that directly affects the firm's production.

The rest of the paper is organized as follows. Section 2 presents the base model and Section 3 analyzes it. Section 5 considers a modified setting with a board of directors and considers a implementation of corporate budgeting practices and Section 6 presents our conclusions.

2 The model

We consider a firm that operates in a risk-neutral economy. The firm is run by an entrepreneur ("the principal") and requires the input from a penniless lower-level manager ("the agent") who is subject to limited liability and a zero reservation wage.[7] The technology of the firm is described by a decreasing returns to scale stochastic production function:

$$Q(e,\theta,k) = z(e,\theta)k - \frac{1}{2}k^2. \tag{1}$$

In particular, this technology combines two inputs, a capital investment k (i.e., the firm's scale) which is subject to a quadratic cost $g(k) \equiv \frac{1}{2}k^2$ and the agent's effort e, which influences the probability of success of the firm $z(e,\theta)$.[8] Specifically, the success probability

$$z(e,\theta) = \begin{cases} r > 0 & \text{with prob. } \theta e \\ 0 & \text{with prob. } (1-\theta e) \end{cases} \tag{2}$$

is determined by (i) an exogenous random shock θ, which is privately observed by the principal,

$$\theta = \begin{cases} 1 & \text{with prob. } (1-\pi) \equiv p_1 \\ \beta > 1 & \text{with prob. } \pi \equiv p_\beta \end{cases}$$

[7] In this section we identify the entrepreneur (the principal) with the firm. In Section 5, however, we separate ownership from control by considering a third party (e.g., a board) with authority on a number of firm's policies (e.g., capital budgeting and financing policy) that affect the incentives of the entrepreneur.

[8] Alternatively, the firm's production function can be described as $H(e,\theta,I) = z(e,\theta)\sqrt{2I} - I$, i.e., the scale of the firm increases by the factor $\sqrt{2I}$ per unit of capital invested.

and (ii) the agent's privately exerted effort $e \in [0, \frac{1}{\beta})$. The agent's cost of effort is given by:

$$h(e,k) = \frac{1}{2}ce^2k,$$

which is quadratic in e (i.e., $\frac{d^2h}{de^2} = ck$) and increases linearly with the firm's scale (i.e., $\frac{dh}{dk} = \frac{1}{2}ce^2$).[9] In what follows, we implicitly assume that c is sufficiently large that the optimal effort choice of the agent lies within the bounds $[0, 1/\beta)$ (so that expression (2) is well defined).

The timing of events is as follows: At $t = 0$, before observing θ, the principal offers the agent a compensation schedule contingent on the firm's scale and the realized output: $w(k,q)$. At $t = 1$, the principal privately observes θ and chooses the scale k. At $t = 2$, the agent makes an unobservable effort choice e. At $t = 3$, firm output q is realized and contracts are settled. Figure 1 summarizes the timing of events.

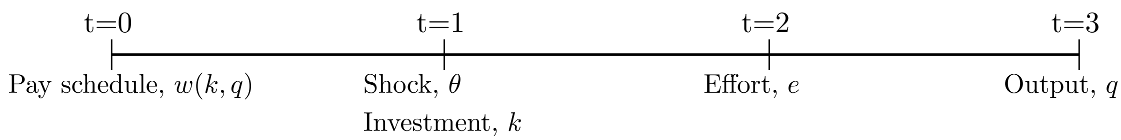

Figure 1: Timing of Events

We analyze the model under the assumption that the principal can commit to any policy at $t = 0$ that is contingent on any observable information. Specifically, the principal can commit to a specific investment choice and offer compensation schedules $w(k,q)$ that are contingent on both the investment k and output, q. However, since the agent's effort is unobservable, the agent's cannot commit to any level of effort in advance and makes his effort choice that is optimal based on the available information (which can include the choice of investment made by the principal).

3 Model analysis

3.1 Observable productivity shock θ

As a benchmark we analyze the case in which both the principal and the agent observe the productivity shock θ and there is still managerial moral hazard (i.e., the agent makes an unobservable effort choice). We denote as $k = \{k_1, k_\beta\}$ and $e = \{e_1, e_\beta\}$ the investment

[9] A setting in which effort costs are independent of the firm scale $h(e) = \frac{1}{2}ce^2$ but in which production costs are cubic on firm scale, i.e., $g(k) \equiv k^3$, while less tractable, produces similar results.

and effort levels that correspond to the two possible realizations of $\theta = \{1, \beta\}$. Without loss of generality, we restrict the analysis to finding compensation contracts that offer non-negative payments when the output is high ($z = r$) and a zero-payment when the output is low ($z = 0$), i.e., $w = \{w_1, w_\beta\}$ when $\theta = \{1, \beta\}$ is realized.[10] Therefore, the principal's problem can be expressed as

$$\max_{k,w,e} \; V = \sum_{\theta=\{1,\beta\}} p_\theta \left[(rk_\theta - w_\theta)\theta e_\theta - \frac{1}{2}k_\theta^2 \right] \tag{3}$$

s.t.

$$e_\theta = \arg\max_e \{w_\theta \theta e - \frac{1}{2}k_\theta c e^2\}, \qquad \text{for } \theta = \{1, \beta\} \tag{4}$$

$$w_\theta \theta e_\theta - \frac{1}{2}k_\theta c e_\theta^2 \geq 0, \qquad \text{for } \theta = \{1, \beta\}. \tag{5}$$

For each realization of the shock, $\theta = \{1, \beta\}$, the principal maximizes firm value (3), subject to the corresponding optimal effort choice (4), and individual rationality constraint (5). The problem can be simplified because the agent's limited liability requires $w_\theta \geq 0$ and, since $e = 0$ is feasible, constraints (5) always hold.

By substituting the first-order condition of (4) in (3) we get:

$$\max_{k,w} \; V = \sum_{\theta=\{1,\beta\}} p_\theta \left[(rk_\theta - w_\theta)\theta \frac{w_\theta \theta}{k_\theta c} - \frac{1}{2}k_\theta^2 \right], \tag{6}$$

the solution of which leads to the following proposition:

Proposition 1 *For $\theta = \{1, \beta\}$ the optimal compensation, investment and effort are:*

$$w_\theta^* = \frac{r^3\theta^2}{8c}, \; k_\theta^* = \frac{r^2\theta^2}{4c} \; \textit{and} \; e_\theta^* = \frac{r\theta}{2c}. \tag{7}$$

From Proposition 1, it follows that optimal compensation, investment and effort are increasing in θ, i.e., $w_\beta^* > w_1^*$, $k_\beta^* > k_1^*$ and $e_\beta^* > e_1^*$. Since effort is more valuable when the firm's prospects are better (i.e., $\theta = \beta$), the optimal policy is to invest more and to induce a greater managerial effort in the more favorable state. A closer examination of the equilibrium relations among the endogenous variables sheds some light on the main underlying economic mechanism of the model. Specifically, effort increases with managerial compensation and decreases with investment ($e_\theta^* = \frac{\theta w_\theta^*}{c k_\theta^*}$), and compensation and investment

[10]In the appendix we show that focusing on these compensation contracts is without loss of generality.

are proportional to each other ($w^*_\theta = \frac{r}{2}k^*_\theta$) which implies that the optimal compensation is a sharing rule that is independent of θ, ($\frac{w^*_\theta}{rk^*_\theta} = \frac{1}{2}$). Finally, for future comparisons, we express firm value in this benchmark case as:

$$V^* = \frac{1}{2}\left[(1-\pi){k^*_1}^2 + \pi {k^*_\beta}^2\right]. \tag{8}$$

3.2 Unobservable productivity shock θ

The contractual arrangement described in Proposition 1 requires that the agent observes θ. If the shock θ is unobservable, the principal may have an incentive to choose the investment and compensation to convey information about θ, which, in turn, affects the agent's effort choice.

Technically, this is a mechanism design problem with an informed principal and an agent who makes an unobservable and privately costly effort choice. In such a framework, the revelation principle does not apply, that is, the optimal mechanism need not be a direct mechanism in which the principal truthfully announces θ and the agent's action is a function of the θ revealed by the principal. This is because, even though the principal can commit to any allocation as a function of the disclosed θ, the agent cannot commit to a specific effort choice since effort is unobservable to anyone but to the agent himself.[11]

In the appendix we show that we can restrict the search for the optimal contract to two investment-compensation pairs, i.e., $\{k_a, w_a\}$ and $\{k_b, w_b\}$. Within this set of contracts three possibilities arise. If the principal's choices depend on the observed type, (i.e., the principal chooses k_a and w_a after observing $\theta = 1$ and $k_b \neq k_a$ or $w_b \neq w_a$ after observing $\theta = \beta$), then those choices communicate his private information to the agent. We refer to this case as *separation*. If, however, $k_a = k_b$ and $w_a = w_b$, then the principal's choices convey no information. We refer to this case as *pooling*. A third possibility would be *partial pooling* where the principal mixes between the two investment-compensation pairs.[12]

In what follows, we search for the optimal contract under separation and pooling and

[11] See Bester and Strausz (2001) for an analysis of the optimal contract in a dynamic adverse selection setting where the principal cannot commit to a specific allocation after the agent's first period choice and Strausz (2009) for a setting in which, similarly, to ours, the revelation principle does not apply due to the unobservability of the agent's effort.

[12] Formally, this situation would require us to define $\{\sigma_a, \sigma_b\}$ as the probabilities that the principal chooses $\{(k_a, w_a), (k_b, w_b)\}$ after observing $\{1, \beta\}$ and to solve for the optimal level of information revelation, i.e, the optimal $\{\sigma_a, \sigma_b\}$.

compare firm value when these contracts are implemented. As shown in the appendix, this is without loss of generality since firm value under partial pooling is lower than firm value when either the optimal separation or the optimal pooling contract is implemented.

3.3 Optimal separation policy

Under separation, the principal's choices of managerial compensation and firm investment depend on the firm's type θ, i.e, $w^s = \{w_1^s, w_\beta^s\}$ and $k^s = \{k_1^s, k_\beta^s\}$. Furthermore, since θ is unobservable to the agent, these choices are subject to incentive compatibility (i.e., "truthtelling") constraints. We define $V_\theta^{\hat{\theta}}$ as the firm value when the principal observes θ and chooses $w_{\hat{\theta}}^s$ and $k_{\hat{\theta}}^s$ for $\hat{\theta}, \theta = \{1, \beta\}$:

$$V_\theta^{\hat{\theta}} \equiv (rk_{\hat{\theta}}^s - w_{\hat{\theta}}^s)\theta e_{\hat{\theta}}^s - \frac{1}{2} k_{\hat{\theta}}^{s^2} \tag{9}$$

where $e_{\hat{\theta}}^s = \arg\max_e \{w_{\hat{\theta}}^s \hat{\theta} e - k_{\hat{\theta}}^s c e^2\}$ for $\hat{\theta} = \{1, \beta\}$ and denote as $e^s = \{e_1^s, e_\beta^s\}$ the corresponding agent's effort's vector. Thus, the principal's problem can be expressed as:

$$\max_{k^s, w^s, e^s} \quad V^s = \sum_{\theta = \{1,\beta\}} p_\theta \left[(rk_\theta^s - w_\theta^s)\theta e_\theta^s - \frac{1}{2} k_\theta^{s^2} \right] \tag{10}$$

s.t.:

$$e_\theta^s = \arg\max_e \{w_\theta^s \theta e - \frac{1}{2} k_\theta^s c e^2\}, \qquad \text{for } \theta = \{1, \beta\} \tag{11}$$

$$w_\theta^s \theta e_\theta^s - \frac{1}{2} k_\theta^s c e_\theta^{s^2} \geq 0, \qquad \text{for } \theta = \{1, \beta\} \tag{12}$$

$$V_\theta^\theta \geq V_\theta^{\hat{\theta}}, \qquad \text{for } \theta, \hat{\theta} = \{1, \beta\} \text{ and } \hat{\theta} \neq \theta. \tag{13}$$

Formally, problem (10)-(13) consists of the addition of the IC constraints (13) to the benchmark problem (3)-(5) where θ is observable. To solve it, we proceed by first ignoring the IC constraint of the high productivity firm, i.e., $V_\beta^\beta \geq V_\beta^1$, and assuming that the IC of the low productivity firm binds, i.e., $V_1^1 = V_1^\beta$.[13] Then, we substitute it into the objective function (10) and derive the corresponding first order conditions to get:[14]

Proposition 2 *In separation, the optimal managerial compensation and investment are:*

$$w^{s*} = \{w_1^*, \Delta w_\beta^*\} \quad \textit{and} \quad k^{s*} = \{k_1^*, \Delta k_\beta^*\} \tag{14}$$

[13]See the appendix for a formal argument that shows that $V_\beta^\beta \geq V_\beta^1$ holds in the optimal solution.

[14]A positive compensation when output is low (i.e., $q = 0$) could, in principle, be part of a separating mechanism. In the appendix we show that in the optimal separating contract this is not the case and that, as in the benchmark case, restricting to zero compensation when the output is zero is without loss of generality.

where $\Delta = \max\left\{\frac{1+\sqrt{1-1/\beta^2}}{\beta}, 1\right\} \geq 1$. *With these choices, the agent's effort is* $e_\theta^{s*} = e_\theta^* = \frac{\theta r}{2c}$ *and the principal's payoff is*

$$V^{s*} = \frac{1}{2}\left[(1-\pi)k_1^{*^2} + \pi\Delta(2-\Delta)k_\beta^{*^2}\right]. \tag{15}$$

A comparison between Propositions 1 and 2 shows that, relative to the case in which θ is observable, compensation and investment remain unaltered for the low prospect firm ($w_1^* = w_1^{s*}$; and $k_1^* = k_1^{s*}$) but increase by the factor Δ ($k_\beta^{s*} = \Delta k_\beta^*$ and $w_\beta^{s*} = \Delta w_\beta^*$) for the high prospect firm. For both types of firms, however, effort equals the effort exerted in the observable θ case ($e_\theta^* = e_\theta^{s*}$) since effort depends on the ratio between compensation and investment ($e_\theta^* = \frac{\theta w_\theta^{s*}}{c k_\theta^{s*}}$) which remains unaltered relative to the effort level in the observable case.

The overinvestment factor Δ summarizes the main intuition of the analysis in the separation case. A high prospect firm must overinvest (i.e., increase its scale) to credibly communicate to the agent the presence of high prospects. Therefore, the optimal policy under separation consists of the determination of Δ, namely the minimum required level of overinvestment that makes such communication credible to the agent.

As Proposition 2 indicates, the overinvestment factor Δ depends on β, which measures the difference in the size of productivity shocks across firms. As a function of β, the overinvestment factor Δ has an inverted U-shape that takes its maximum at $\hat{\beta} = \frac{2\sqrt{3}}{3} \cong 1.15$ and its minimum (i.e., $\Delta = 1$) when $\beta \geq \beta^* \cong 1.84$. In other words, when differences in productivity are not large (i.e., when $\beta < \beta^*$) the high prospect firm overinvests to convey its type to the agent. However, when the differences in productivity are large enough (i.e., when $\beta \geq \beta^*$) there is no overinvestment in the optimal separation policy and the solution under separation corresponds to the solution in the benchmark case when the agent cannot observe θ.[15]

It is worth noting that overinvestment distortions emerge as a component of the optimal separation policy even if there are, in principle, other ways in which a separating mechanism could be designed. For example, rather than distorting investment the firm could commit to use a distorted compensation policy to convey information about its type to the agent. Overinvestment distortions are particularly effective, however, since a type 1 firm finds it

[15] See the appendix for more details on the relation between Δ and β.

costlier to overinvest than a type β firm because, given its lower expected productivity, the marginal benefits of a high investment level are lower for the low type firm.[16] In contrast, distorted compensation is not effective because, for incentive reasons, the compensation needs to be contingent on success and the probability of success is higher when $\theta = \beta$. In these conditions, an upward shift in w_θ^s (i.e., an increase of w_1^s and w_β^s by a constant), does not work because it is costlier for a type β than for a type 1 firm.[17] Alternatively, the firm could commit to burn money (i.e., engage in wealth destruction regardless of the firm scale or output success) as part of the separation mechanism. Such a device, however, is also inferior when compared with overinvestment. This is so because, in contrast to the overinvestment distortion which is relatively more costly for the type 1, money burning is equally costly to both firm types.

3.4 Optimal pooling policy

Rather than a separation policy, the principal can follow a pooling policy in which he commits (before observing θ) to make investment and compensation choices that are independent of the observed θ.[18] Under pooling, the principal avoids the disclosure of his information on θ and thus the agent also makes an effort choice which is independent of θ. We refer to $\bar{k}$ and $\bar{w}$ as the investment and the compensation in case of success and to $\bar{\theta} \equiv \pi\beta + (1-\pi)1$ as the average productivity shock. To find the optimal pooling policy, the principal solves:

$$\max_{\bar{k},\bar{w},\bar{e}} \; V_p = (r\bar{k} - \bar{w})\bar{\theta}\bar{e} - \frac{1}{2}\bar{k}^2 \tag{16}$$

s.t.:

$$\bar{e} = \arg\max_{e}\{\bar{w}\bar{\theta}e - \frac{1}{2}\bar{k}ce^2\}. \tag{17}$$

Since the principal's choices do not convey any information, the agent makes his effort decision based on the average firm type. Thus, the solution to the problem is given by:

[16] Since agent's effort costs increase with the firm's scale, the overinvestment distortion is accompanied by a corresponding distortion in compensation in order to restore the proper effort incentives to the agent.

[17] In the appendix we provide a more extensive analytical argument to show that other distortions in compensation (e.g., increasing compensation in case of failure) are suboptimal as well.

[18] Technically this commitment would require that the principal set self-imposed penalties in the optimal mechanism in case that he chooses investment or compensation actions different from those considering in the pooling policy under consideration.

Proposition 3 *The optimal compensation and firm investment under pooling are:*

$$\bar{w}^* = \frac{r^3\bar{\theta}^2}{8c} \quad \text{and} \quad \bar{k}^* = \frac{r^2\bar{\theta}^2}{4c}. \tag{18}$$

With these choices, the agent's effort is $\bar{e}^* = \frac{r\bar{\theta}}{2c}$ *and the principal's payoff is* $V_p^* = \frac{1}{2}\bar{k}^{*2}$.

Under pooling, the agent's effort is the average of the two efforts levels under the benchmark case in which θ is observable. Also, compensation is a sharing rule as it was in the benchmark and separation cases i.e., $\frac{\bar{w}^*}{r\bar{k}^*} = \frac{1}{2}$. In addition, the firm's investment choice does not depend on the realized θ and, thus, will exhibit overinvestment when the realized type is 1 and underinvestment when the realized type is β.

Intuitively, the pooling case illustrates how rigidities in investment and compensation can be part of an optimal capital budgeting policy and why a firm may find it valuable to commit to ignore information when it makes investment decisions. More specifically, the results illustrate why firms might commit to a maximum investment level. This interpretation of pooling as an investment limit is consistent with a number of studies that show that firms may exhibit a tendency toward capital rationing when they allocate capital in different business units (e.g., Ross 1986 documents that fixed capital budgets are commonly employed by firms).

3.5 Separation or pooling?

Compared with the separation case, with pooling the firm does not overinvest or overcompensate managers when $\theta = \beta$. The cost, however, is that the firm does not tailor its investment expenditures to its realized marginal productivity. The optimal choice between pooling and separation depends on the trade-off between these costs and benefits as the next proposition states.

Proposition 4 *(Comparative statics) Separation is more likely to be the optimal policy: (i) the lower the likelihood of a high productivity firm (i.e., lower* π*) and (ii) the larger the difference in productivity among firm types (i.e., higher* β*).*

Figure 2 displays the regions in the space (β, π) where each policy is optimal. As Figure 1 shows, pooling is locally optimal, that is, when productivity levels are sufficiently close (i.e., $\beta \to 1$) pooling emerges as the optimal investment policy. To the extent that firms can exhibit a wide range of productivity levels, this observation suggests that investment rigidities (i.e., pre-specified discrete levels of investment) are likely to be part of any optimal capital budgeting policy. This observation will have important implications when we formally examine the case where there are a continuum of types.

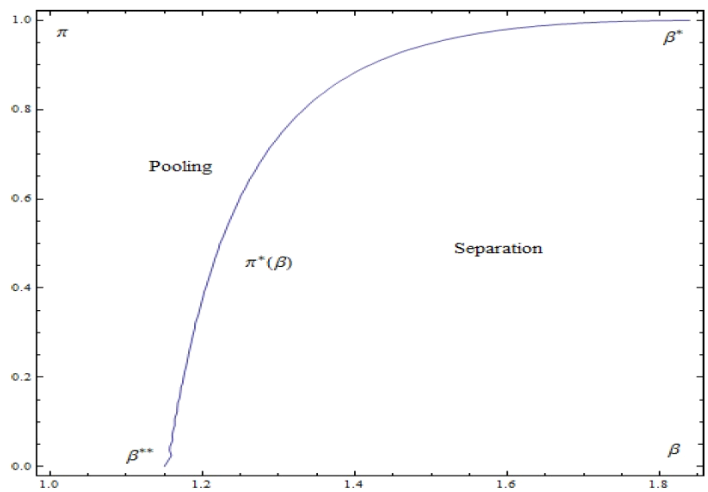

Figure 2: Comparative Statics

To gain some intuition on the previous comparative statics results it is useful to consider some limiting cases. To illustrate the negative effect of π (the probability of a type β firm) on the likelihood of separation we consider the case in which π is close to 1 (i.e., a β type firm is highly likely). In this case, for a type β firm, the distortion is large under separation (the expected overinvestment cost is large) and small under pooling (the investment is close to the type β full-information investment level).[19]

The positive effect of β (the difference in productivity between firm types) on the likelihood of separation can be easily shown when β is large enough and close to β^*. In this case, the efficiency loss due to pooling is large because the average investment is far from the full information level for both types. By contrast, the separation inefficiency (namely type β overinvesment) diminishes when β is sufficiently close to β^*. In fact, as β moves toward β^* the optimal separation contract converges to the first best contract with efficient

[19]For type 1 firms, the investment inefficiency are large under pooling and nil under separation. However, when $\pi \to 1$ these effects are of second order importance relative to the trade-off that arises for a type β firm.

investment for the type 1 firm and an overinvestment level that approaches zero for the type β firm.

4 A continuum of types

In this section we consider the design of the optimal mechanism in a setting where the productivity shock θ can take any value in the interval $[1, \bar{\beta}]$. Considering a continuous type space allows us to check the robustness of our previous findings and to better investigate the idea of investment rigidities. Formally we assume the timing of the model is as before (see Figure 1) but that rather than a two-type distribution the (common) prior distribution of θ is described by the density $f(\theta) > 0$ (with $|f'| \leq M$ and cumulative distribution $F(\theta)$) for $\theta \in [1, \bar{\beta}]$.

We analyze two aspects of the problem. First, we describe the optimal mechanism among those in which different types of firms choose different firm policies (separation). Second, we consider the design of the optimal mechanism among those that allow for the possibility of type bunching. In both cases, since the continuous type model presents a number of technical difficulties, we describe the main results of the analysis in the text and leave most technical derivations for the appendix.

4.1 Optimal separation policy

Under separation, the objective is to find a schedule of non-negative pairs $\{(w^s_\theta, k^s_\theta)\}_{\theta \in [1,\bar{\beta}]}$ such that (i) firms of different types choose different managerial compensation (i.e., bonus) and investment (i.e., $(w^s_{\theta'}, k^s_{\theta'}) \neq (w^s_{\theta''}, k^s_{\theta''})$ if $\theta' \neq \theta''$) and (ii) the schedule of pairs $\{(w^s_\theta, k^s_\theta)\}_{\theta \in [1,\bar{\beta}]}$ maximizes expected firm value. We refer to w^s_θ and k^s_θ as the compensation and investment choices of a type θ firm, and to e^s_θ as the agent's induced level of effort under the firm's choices.[20]

Since the agent cannot observe θ, the firm's choices within $\{(w^s_\theta, k^s_\theta)\}_{\theta \in [1,\bar{\beta}]}$ are subject to incentive compatibility constraints (a type θ firm must find it optimal to choose (w^s_θ, k^s_θ) rather than any other feasible pair $(w^s_{\hat{\theta}}, k^s_{\hat{\theta}})$).[21] Specifically, we define $V^{\hat{\theta}}_\theta$ as the value of a

[20] As in the two type case, we show in the appendix that optimal managerial compensation is set at zero when the realized output is zero ($z = 0$) so that it is completely determined by w^s_θ, a non-negative bonus paid after the positive output is realized (i.e., when $z = r$).

[21] Without loss of generality, the search for the optimal separation mechanism restricts the firm's choices

type θ firm that chooses the compensation and investment of a type $\hat{\theta}$ firm ($w^s_{\hat{\theta}}$ and $k^s_{\hat{\theta}}$):

$$V^{\hat{\theta}}_{\theta} \equiv (rk^s_{\hat{\theta}} - w^s_{\hat{\theta}})\theta e^s_{\hat{\theta}} - \frac{1}{2}k^{s2}_{\hat{\theta}} \tag{19}$$

where $e^s_{\hat{\theta}} = \arg\max_e\{w^s_{\hat{\theta}}\hat{\theta}e - k^s_{\hat{\theta}}ce^2\}$ for $\hat{\theta} \in [1, \bar{\beta}]$. We denote as $V^s_\theta \equiv V^\theta_\theta$ the type θ firm value when the firm chooses (w^s_θ, k^s_θ), i.e., the pair in the schedule designated for its own type. Thus, the principal's problem can be expressed as:

$$\max_{\{w^s_\theta, k^s_\theta, e^s_\theta\}_{\theta\in[1,\bar{\beta}]}} \int_1^{\bar{\beta}} V^s_\theta f(\theta)d\theta \tag{20}$$

s.t.:

$$e^s_\theta = \arg\max_e\{w^s_\theta\theta e - \tfrac{1}{2}k^s_\theta ce^2\} \qquad \text{for } \theta \in [1, \bar{\beta}], \tag{21}$$

$$w^s_\theta\theta e^s_\theta - \tfrac{1}{2}k^s_\theta c{e^s_\theta}^2 \geq 0 \qquad \text{for } \theta \in [1, \bar{\beta}], \tag{22}$$

$$V^s_\theta \geq V^{\hat{\theta}}_\theta \qquad \text{for } \theta, \hat{\theta} \in [1, \bar{\beta}] \text{ and } \theta \neq \hat{\theta}, \tag{23}$$

$$V^s_\theta \geq 0 \qquad \text{for } \theta \in [1, \bar{\beta}]. \tag{24}$$

The following proposition characterizes the optimal mechanism under separation.

Proposition 5 *In the optimal separating mechanism ($\{(w^{*s}_\theta, k^{*s}_\theta)\}_{\theta\in[1,\bar{\beta}]}$) the optimal investment k^{s*}_θ is defined by: $8c{k^{s*}_\theta}^3 - 3\theta^2r^2{k^{s*}_\theta}^2 + 4c{k^{s*}_1}^3 = 0$. Furthermore, the optimal compensation is proportional to investment, $w^{s*}_\theta = \frac{r}{2}k^{s*}_\theta$, and the optimal effort is given by $e^{s*}_\theta = \frac{r\theta}{2c}$.*

As the above proposition indicates, a number of features of the optimal mechanism with a continuum of types resemble those in the optimal mechanism with two types. First, the separating mechanism is characterized by overinvestment. Specifically, relative to a setting in which firm type is observable, all but the lowest type overinvest and the lowest type does not distort investment (i.e., $k^{s*}_\theta > k^*_\theta$ for $\theta > 1$ and $k^{s*}_1 = k^*_1$). Second, optimal compensation is a sharing rule (i.e., the ratio of compensation over output is constant, $\frac{w^{s*}_\theta}{rk^{s*}_\theta} = \frac{1}{2}$). Third, effort is not distorted relative to the case of observable types and, as in the case with observable types, depends on the ratio of investment over compensation. While type unobservability leads to overinvestment and to overcompensation for each type,

after observing θ to those within the schedule $\{(w^s_\theta, k^s_\theta)\}_{\theta\in[1,\bar{\beta}]}$. Such a restriction can be achieved if the firm commits (before observing θ) to self-impose a large penalty (e.g., a large wealth transfer) in case it chooses a pair outside the schedule $\{(w^s_\theta, k^s_\theta)\}_{\theta\in[1,\bar{\beta}]}$.

investment and compensation increase by the same factor and thus their ratio remains unaltered.[22]

As we show next, in this setting the optimal mechanism is never characterized by complete separation. However, in a slightly different setting, where the timing of events is a little modified the equilibrium is in fact characterized by complete separation. Specifically in an alternative setting in which the principal is informed about the firm's type *prior* to considering the design of any specific mechanism, and in which the principal can choose firm investment and managerial compensation to signal its type (in order to influence the agent's effort), it can be shown that a Cho and Kreps' (1987) refinement leads to a separating equilibrium that is equivalent to the separating mechanism described in Proposition 5.

4.2 The optimal mechanism

Unlike in the two type case in which the alternative to separation was pooling, with a continuum of types there are many ways in which types can be (partially or fully) pooled together. Indeed, as discussed in Laffont and Tirole (1988) solving for the optimal mechanism with a continuum of types is technically challenging. However in the current setting, as shown in the appendix, we can focus the search for the optimal mechanism within the class of partition mechanisms which we define next.

A partition mechanism Ψ^m is a mechanism defined by a partition $\Psi = \{\varphi_1, \varphi_2, ...\}$ of the type space $[1, \bar{\beta}]$ and by a restriction on the (compensation and investment) policies followed by the firms whose types belong to a given subinterval $\varphi_i \in \Psi$ (i.e., $\{(w_\theta, k_\theta)\}_{\theta \in \varphi_i}$).[23] Specifically, within each φ_i firms' policies are restricted to be either (i) *pooling*, i.e., all firms have equal compensation and investment policies ($(w_{\theta'}, k_{\theta'}) = (w_{\theta''}, k_{\theta''})$ if $\theta', \theta'' \in \varphi_i$ and $\theta' \neq \theta''$) or (ii) *separation*, i.e., all firms have different policies ($(w_{\theta'}, k_{\theta'}) \neq (w_{\theta''}, k_{\theta''})$ if $\theta', \theta'' \in \varphi_i$ and $\theta' \neq \theta''$). We refer to φ_i^p (respectively φ_i^s) as a subinterval in which firms follow a pooling (respectively separation) policy.

Solving for the optimal Ψ^m amounts to identify the partition Ψ and the schedule of non-negative pairs $\{(w_\theta^g, k_\theta^g)\}_{\theta \in [1, \bar{\beta}]}$ associated with Ψ that maximizes expected firm value. We

[22]Formally, $e_\theta^{s*} = \frac{w_\theta^{s*}\theta}{ck_\theta^{s*}} = \frac{r\theta}{2c} = \frac{w_\theta^{*}\theta}{ck_\theta^{*}} = e_\theta^{*}$.

[23]Formally the set $\Psi = \{\varphi_1, \varphi_2, ...\}$ is a partition of $[1, \bar{\beta}]$ if *(i)* $\underset{i \in \mathbb{N}}{\cup} \varphi_i = [1, \bar{\beta}]$, *(ii)* $\varphi_i \cap \varphi_j = \varnothing$, and *(iii)* $\varphi_i = (\underline{\theta}_i, \overline{\theta}_i]$. Notice that the number of subintervals φ_i can be countably infinite.

denote by (w^s_θ, k^s_θ) the compensation and investment choices when $\theta \in \varphi^s_i$ and by $(\bar{w}_{\varphi_i}, \bar{k}_{\varphi_i})$ as the choices when $\theta \in \varphi^p_i$. We refer to e^s_θ and $\bar{e}_{\varphi_i}$ as the corresponding associated efforts and define firm value of a type θ firm when choosing the pair of a type $\hat{\theta}$ firm as:

$$V^{\hat{\theta}}_\theta \equiv \begin{cases} (k^s_{\hat{\theta}} r - w^s_{\hat{\theta}})\theta e^s_{\hat{\theta}} - \frac{1}{2} k^{s^2}_{\hat{\theta}} & \text{if } \hat{\theta} \in \varphi^s_i \\ (\bar{k}_{\varphi_i} r - \bar{w}_{\varphi_i})\theta \bar{e}_{\varphi_i} - \frac{1}{2}\bar{k}^2_{\varphi_i} & \text{if } \hat{\theta} \in \varphi^p_i. \end{cases} \tag{25}$$

Thus, the principal's problem can be expressed as:

$$\max_{\Psi, \{w^g_\theta, k^g_\theta, e^g_\theta\}_{\theta \in [1,\bar{\beta}]}} E[V^\theta_\theta] \tag{26}$$

s.t.:

$$e^s_\theta = \arg\max_e \{w^s_\theta \theta e - \tfrac{1}{2} k^s_\theta c e^2\} \qquad \text{if } \theta \in \varphi^s_i \tag{27}$$

$$\bar{e}_{\varphi_i} = \arg\max_e \{\bar{w}_{\varphi_i} E[\theta|\varphi^p_i] e - \tfrac{1}{2}\bar{k}_{\varphi_i} c e^2\} \qquad \text{if } \hat{\theta} \in \varphi^p_i \tag{28}$$

$$V^\theta_\theta \geq V^{\hat{\theta}}_\theta, \qquad \text{for } \theta, \hat{\theta} \in [1, \bar{\beta}] \text{ and } \theta \neq \hat{\theta}. \tag{29}$$

While we cannot provide a complete characterization of the solution, the following proposition specifies four properties of the optimal mechanism.

Proposition 6 *The optimal mechanism, $\Psi^{m^*} = \{\Psi^*, \{(w^{*g}_\theta, k^{*g}_\theta)\}_{\theta \in [1,\bar{\beta}]}\}$, features the following properties:*

(1) Compensation is proportional to investment, i.e., $w^{*g}_\theta = \frac{r}{2} k^{*g}_\theta$.

(2) Investment is weakly increasing in firm productivity θ, i.e., if $\theta' < \theta''$ then $k^{*g}_{\theta'} \leq k^{*g}_{\theta''}$.

(3) Pooling holds at the top and bottom of the type distribution, i.e., $\exists\ \theta_L \leq \theta_H$ such that $\{(w^{*g}_\theta, k^{*g}_\theta)\}_{\theta \in [1,\bar{\beta}]} \supset \{(\bar{w}^*_L, \bar{k}^*_L)\}_{\theta \in [1,\theta_L]}$ and $\{(\bar{w}^*_H, \bar{k}^*_H)\}_{\theta \in [\theta_H, \bar{\theta}]}$.

(4) There is overinvestment in every subinterval of Ψ^*, i.e., $k^{*s}_\theta > k^*_\theta$ if $\theta \in \varphi^{s*}_i$ and $\bar{k}^*_{\varphi_i} > \arg\max_k \{E[V^\theta_\theta | \varphi^{p*}_i]\}$ if $\theta \in \varphi^{p*}_i$.

The previous proposition describes four features of Ψ^{m^*} that have relevant economic implications. First, as in the two-type model, compensation is proportional to investment (and hence to output) which implies that compensation is a *sharing rule* (i.e., $\frac{w^{*g}_\theta}{r k^{*g}_\theta} = \frac{1}{2}$). Technically, this property simplifies the description of the optimal schedule which is fully characterized by the relationship between investment and firm's type i.e., k^{*g}_θ. Second, firm's investment is *monotonically* increasing in its productivity; that is, more productive firms invest at least as much as their less productive counterparts. Third, Ψ^{m^*} features pooling at the top and bottom of the type distribution. Intuitively, this implies that the principal commits to minimum and maximum *investment limits* in order to implement the optimal capital budgeting policy.

Top and bottom pooling in Ψ^{m^*} is consistent with Proposition 4 which states that pooling is locally optimal in the two-type model. This observation suggests the optimal mechanism is likely to be characterized by a partition of non-overlapping pooling intervals. We have been unable, however, to confirm this conjecture. In particular, while we can construct cases of pooling with one or two adjacent intervals we cannot formally exclude the possibility of separation in some middle intervals of the partition.[24]

Proposition 6 also states that overinvestment is pervasive in the optimal mechanism. Specifically, it states that in separating intervals all firm-types overinvest and that firms overinvest on average in pooling intervals. This pervasive overinvestment is in contrast with the typical underinvestment result obtained in models that focus on managerial private benefits of investment (e.g., Bernardo, A., H. Cai, and J. Luo, 2001). Intuitively, this difference occurs because, in our setting, overinvestment is costly to the principal but it is a relatively efficient way of inducing type separation (i.e., of conveying private information) to the agent.

We conclude this section by stating a corollary that combines intuition on investment limits and overinvestment results as stated in Proposition 6:

Corollary 1 *The minimum investment limit is above the efficient level of investment of the lowest productivity firm, i.e.,* $\bar{k}_L^* > k_1^*$.

The corollary follows because the bottom interval in the optimal partition is pooling and because, on average, firms overinvest in pooling intervals.

5 Corporate capital budgeting

In this section we consider a number of issues observed in the practice of capital budgeting and investigate their effects in a setting like ours. In particular, we consider three specific

[24]For example, numerical analysis (available upon request) shows that a two-pooling-interval partition is optimal if θ is uniformly distributed on $[1, 2]$.

questions: the use of debt financing, the implementation of a distorted discount rate policy and other issues relating to compensation.

To analyze these questions, we go back to our simpler two type model and extend it by adding a third party. Specifically, in addition to the entrepreneur, who obtains information, and the employee who exerts effort, we introduce a third party that provides debt financing to fund the project or sets capital budgeting policies. Depending on the question that we consider, this party is interpreted as an external financier or a board of directors with authority to impose rules on the entrepreneur.

5.1 A role for debt financing

We relax the assumption that the entrepreneur uses his own wealth to fund the project and consider the possibility that debt finance can be also raised. Specifically we assume that, before starting operations, the firm can issue debt with face value d in a competitive capital market where the required rate of return is (normalized to) zero. We consider the following timing of events:

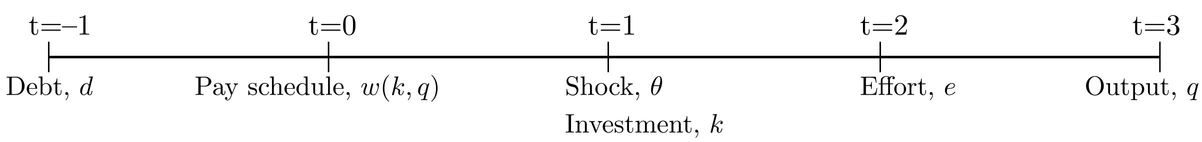

Figure 3: Timing of Events

As the previous figure indicates, we assume that debt financing is obtained before the project starts (at $t = -1$) and, therefore, that there is no asymmetry of information between the entrepreneur and the debtholders. We first consider the case of separation and then briefly discuss the effects of debt financing on the pooling policy.

5.1.1 Debt and separation policy

To analyze debt issuances under a separation policy we first examine the effect of debt on other choice variables and then consider the optimal debt choice. For a given debt level d, we denote by $k^d = \{k^d_1, k^d_\beta\}$, $w^d = \{w^d_1, w^d_\beta\}$ and $e^d = \{e^d_1, e^d_\beta\}$ the investment, bonus and effort exerted when $\theta = \{1, \beta\}$ and by $V^{\hat{\theta}}_{\theta,d} \equiv (rk^d_{\hat{\theta}} - w^d_{\hat{\theta}} - d)\theta e^d_{\hat{\theta}} - \frac{1}{2}k^{d^2}_{\theta}$ the principal's payoff when, after observing θ, offers $w^d_{\hat{\theta}}$ and invests $k^d_{\hat{\theta}}$ for $\hat{\theta}, \theta = \{1, \beta\}$. As the equityholder of the firm, the principal solves at $t = 1$:

$$\max_{k^d, w^d, e^d} \quad V^d = \sum_{\theta=\{1,\beta\}} p_\theta[(rk^d_\theta - w^d_\theta - d)\theta e^d_\theta - \frac{1}{2}k^{d^2}_\theta] \tag{30}$$

s.t.:

$$e^d_\theta = \arg\max_e \{w^d_\theta \theta e - \frac{1}{2}ce^2 k^d_\theta\} \qquad \text{for } \theta = \{1, \beta\} \tag{31}$$

$$V^\theta_{\theta,d} \geq V^{\hat{\theta}}_{\theta,d} \qquad \text{for } \theta, \hat{\theta} = \{1, \beta\} \text{ and } \hat{\theta} \neq \theta. \tag{32}$$

The solution to the previous problem can be characterized as follows:

Lemma 1 *Leverage reduces the wage, effort, and investment relative to the unlevered case, i.e., $w^{d*}_\theta < w^{s*}_\theta$, $e^{d*}_\theta < e^{s*}_\theta$, and $k^{d*}_\theta < k^{s*}_\theta$ for $\theta = \{1, \beta\}$.*

As stated in Lemma 1 the use of debt financing reduces investment for all firm's types. This result is easily understood in the benchmark case in which θ is publicly observed. In such a case, the principal of a levered firm would choose an investment level lower than k^*_θ, the investment level of the unlevered firm. This effect occurs because the principal of the levered firm internalizes the cost of the marginal unit invested (which comes from his own funds) but internalizes only part of the expected revenue (which goes partly to the debtholders). A similar effect occurs on effort (whose positive effect is shared with the debtholders) and hence on wages.

Taking into account the effects of leverage described in Lemma 1, at $t = -1$, the principal chooses leverage to maximize firm value. Formally the principal solves the following problem at $t = -1$:[25]

$$\max_d V^d = \sum_{\theta=\{1,\beta\}} p_\theta[(rk^{d*}_\theta - w^{d*}_\theta)\theta e^{d*}_\theta - \frac{1}{2}k^{d*^2}_\theta]. \tag{33}$$

The solution to the principal's problem is described in the following proposition.

Proposition 7 *If without leverage the optimal firm policy is separation with overinvestment (i.e., $\beta < \beta^*$) then leverage increases firm value ($d^* > 0$); if it is separation with efficient investment ($\beta \geq \beta^*$) then the optimal amount of leverage is zero ($d^* = 0$).*

[25]Since the debt market is competitive and debt is fairly priced, at the time of debt issuance maximizing firm value is equivalent to maximize equity value.

This proposition states that debt can create value when θ is unobservable under separation. This occurs because leverage creates debt overhang that affects the IC constraints (32) which, in this setting, has the positive effect of reducing the incentives of low quality firms to mimic high quality ones. Therefore, leverage increases firm value when the positive incentive effects of debt overhang offset the negative effects of debt overhang due to investment distortions. As stated in Proposition 7, when separation requires overinvestment, the gain obtained by reducing overinvestment for the high quality firm outweighs the cost due to the underinvestment created on the low quality firm.[26] Of course this is not the case in the case of separation without overinvestment ($\beta \geq \beta^*$). There, the use of debt just creates underinvestment in both types and leads to a reduction in firm value.

5.1.2 Debt and pooling policy

Under a pooling policy, the firm commits at $t = 0$ to a fixed level of investment and compensation which eliminates the possibility to mimic other types' choices in order to influence the agent's effort. As a result, under pooling, leverage has no positive effects on "mimicking" incentives and can merely create a debt overhang distortion. Henceforth, the following proposition holds:

Proposition 8 *If pooling is the optimal firm policy, then the optimal amount of debt to fund the project is zero ($d^* = 0$).*

From the analysis of endogenous leverage we can make several observations. First, the combination of Propositions 7 and 8 implies that when leverage is endogenous, separation is more likely to be the optimal firm policy. Second, the analysis identifies a positive effect of leverage which has not been considered in the literature so far. In situations in which firms overinvest to convey information about their prospects, leverage affects the transmission of information (i.e., the incentives of low prospect firms to mimic their high prospects counterparts) and can create firm value. It is noteworthy that, in contrast to other positive effects of debt which are related to tax distortions or to the need to correct managerial

[26] Intuitively, this is a manifestation of the envelope theorem: for small levels of leverage the reduction in overinvestment (chosen to disuade mimicking from low types) leads to a first order increase in firm value while the underinvestment problem induced in the low quality firm leads to a second order reduction in firm value since it is a (small) reduction from the optimal level of investment.

misincentives, the positive effect of debt occurs in a setting without taxes and in which firm decisions are made in the best interest of the shareholders.

In the analysis, we have focused on the case in which a firm borrows at *t=–1* before knowing its type rather than at *t=1* after knowing it. This modeling choice, which corresponds to a setting in which firms secure external finance before specific investment projects arise, allow us to illustrate in a simple way how a firm's leverage may play a positive role in capital budgeting when its investment expenditures convey information. Furthermore, this choice avoids the technical complications that arise when firm financing is made under asymmetric information and financing choices could convey information on firm's propects. If firms can borrow at *t=1*, then in our setting, a low prospect firm would find it attractive to get funds at the terms offered to the high prospect firm.[27] This suggests that, even in this case, overinvestment rather than leverage is likely to be prevalent in the optimal separating firm policy.

5.2 A role for the board of directors

We now extend the model to consider the interplay of three parties in the capital budgeting process. In addition to the entrepreneur (to whom we now refer to as a "CEO") who obtains information and determines investment expenditures, and the employee who exerts effort, we introduce a third party (i.e., a "board") that sets specific capital budgeting policies and offers compensation contracts to the CEO. One interpretation is that the third party is a venture capitalist that provides funding for the firm and retains some control of its operations. Another is that the informed party is the manager of the division of a conglomerate and the third party includes executives at the firm's headquarters or are members of the board. In either interpretation, the third party is an entity that has the authority to set the firm policy.

The following figure considers the timing of the three-party model:

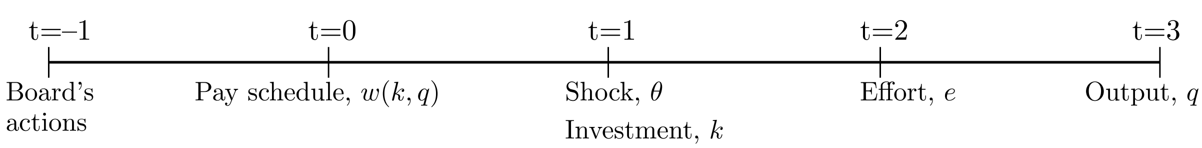

Figure 3: Timing of Events

[27] Since the low type firm has a lower probability of repayment it would find risky debt underpriced.

As shown in Figure 3, the three-party model adds period $t = -1$ where the board sets the policy that determines the CEO's payoffs (i.e., the principal's objective function).

5.2.1 Optimal discount rate policy

We consider first the case where at $t = -1$, the board chooses the discount rate, i, or equivalently, a discount factor, $\delta \equiv \frac{1}{1+i}$ that the CEO uses to evaluate the firm's investment choices. We focus this analysis on the case where separation is the optimal firm policy, and consider the case of pooling briefly below. As a matter of practical implementation, we assume that the board can offer the CEO a compensation contract with "EVA-like" features that induces the CEO to evaluate investment with the discount factor δ.[28] Specifically, the board will offer the CEO a share, α, of the firm's net value at $t = 3$, where net firm value is calculated as revenues minus wages and investment expenditures divided by the factor δ. This implies that if $\delta > 1$ the board encourages the use of capital, relative to the full information case, and if $\delta < 1$ the board discourages the use of capital.

Consider the problem encountered by the CEO at $t = 0$ who must discount the investment's future cash-flows by the factor δ imposed by the board. We denote by $k^\delta = \{k_1^\delta, k_\beta^\delta\}$, $e^\delta = \{e_1^\delta, e_\beta^\delta\}$ and $w^\delta = \{w_1^\delta, w_\beta^\delta\}$ the investment, bonus and effort exerted after $\theta = \{1, \beta\}$ respectively. In addition, we refer to $V_{\theta,\delta}^{\hat{\theta}}$ as the CEO's payoff when, after observing θ, offers $w_{\hat{\theta}}^\delta$ and invests $k_{\hat{\theta}}^\delta$ for $\hat{\theta}, \theta = \{1, \beta\}$:

$$V_{\theta,\delta}^{\hat{\theta}} \equiv (rk_{\hat{\theta}}^\delta - w_{\hat{\theta}}^\delta)\theta e_{\hat{\theta}}^\delta - \frac{1}{2\delta} k_\theta^{\delta^2}. \tag{34}$$

Formally, the CEO solves the following maximization problem:

$$\max_{k^\delta, w^\delta, e^\delta} \quad V_\delta = \sum_{\theta=\{1,\beta\}} p_\theta[(rk_\theta^\delta - w_\theta^\delta)\theta e_\theta^\delta - \frac{1}{2\delta} k_\theta^{\delta^2}] \tag{35}$$

s.t.:

$$e_\theta^\delta = \arg\max_e \{w_\theta^\delta \theta e - \frac{1}{2} k_\theta^\delta c e^2\} \qquad \text{for } \theta = \{1, \beta\} \tag{36}$$

$$V_{\theta,\delta}^{\theta} \geq V_{\theta,\delta}^{\hat{\theta}} \qquad \text{for } \theta, \hat{\theta} = \{1, \beta\} \text{ and } \hat{\theta} \neq \theta. \tag{37}$$

Relative to the case without the board, the agent's problem remains unchanged and the CEO faces the same problem as the principal in the basic model except that the investment

[28] We discuss more general CEO compensation schemes in the next section.

costs are given by $\frac{1}{2\delta}k_\theta^{\delta 2}$ rather than $\frac{1}{2}k_\theta^{\delta 2}$. (Notice that since $\alpha > 0$ is a constant it cancels out in the previous problem, i.e., it does not have any effect on the CEO's choices.) Following the analysis in the last section the solution can be described as follows:

Lemma 2 *The solution of the CEO problem with separation scales investment and wage by δ, respectively ($k_\theta^{\delta *} = \delta k_\theta^{s*}$, $w_\theta^{\delta *} = \delta w_\theta^{s*}$) and leaves the induced agent's effort unaffected ($e_\theta^{\delta *} = e_\theta^{s*}$).*

At $t = -1$ the board takes into account the distortion that the discount factor δ produces on the CEO's behavior at $t = 0$ and, consequently, chooses δ to maximize firm value:

$$\max_\delta V^\delta = \sum_{\theta=\{1,\beta\}} p_\theta[(rk_\theta^{\delta *} - w_\theta^{\delta *})\theta e_\theta^{\delta *} - \frac{1}{2}k_\theta^{\delta * 2}]. \tag{38}$$

The solution to the board's problem is described in the following proposition.

Proposition 9 *If separation with overinvestment characterizes the CEO's investment policy ($\beta < \beta^*$) then the board distorts the discount rate upwards ($\delta^* < 1$). If separation requires no overinvestment ($\beta \geq \beta^*$) then the board does not distort the discount rate ($\delta^* = 1$).*

To better explain the intuition behind the previous proposition it is useful to notice that in problem (38) the board maximizes firm value (i.e., the value determined by discounting cash flows with $\delta = 1$, the "correct" discount factor) by imposing $\delta < 1$ on the CEO when $\beta < \beta^*$. By maximizing a distorted measure of firm value, one in which cashflows are discounted by $\delta < 1$, the CEO reduces the firm's investment expenditures. Intuitively, the benefits of distorting δ can be explained as follows. A lower δ reduces the incentives of a CEO who observes $\theta = 1$ to mimic the behavior of a CEO who observes $\theta = \beta$. A relaxed IC constraint for $\theta = 1$ reduces β's cost to overinvest but creates some underinvestment in $\theta = 1$. In particular, similarly to the case of debt analyzed above, an infinitesimal reduction from $\delta = 1$ simultaneously diminishes w_θ^δ, and k_θ^δ. Since compensation and investment for the low type (w_1^δ and k_1^δ) are at their optimal levels, such a reduction has a second order effect on firm value, while the reduction for the high type (a lower w_β^δ and k_β^δ), which are set at local optima, has a first order effect on firm value. (When separation entails no distortion, $\beta \geq \beta^*$, the optimal discount factor is $\delta = 1$, which keeps the investment and compensation policies undistorted.)

The ability to set a discount rate does not increase firm value in the case of pooling. In this case, if the board sets $\delta \neq 1$, the CEO distorts his choices away from the optimal pooling choices (i.e., $\bar{k}^*$and $\bar{w}^*$). This reduces firm value since, by construction, the optimal pooling choice maximizes firm value relative to the set of choices that convey no information to the agent.[29] For this reason, separation is more likely to be optimal when the hurdle rate is set by the board since a modified hurdle rate may improve the separation policy but does not help the pooling policy.

It should be emphasized that the board creates value by altering the CEO's objective function which, in turn, affects the set of CEO's actions that are incentive compatible for an observed type θ. Without the board, the CEO can also take actions that affect his ex-post investment incentives. If lowering investment (and making it incentive compatible) were the objective, the CEO could do so by committing to transfer output to the agent (or to burn money) when his investment expenditures are high. However, in the two-party case these commitments by the CEO are costly and (as the analysis shows) value-reducing. In contrast, the presence of a board able to set capital budgeting rules changes matters. This is so, because the board acts as the residual claimant (hence neither additional compensation to workers is required nor there is any need to engage in "money burning") and therefore the CEO can effectively credibly commit the information through the investment choices in a more efficient manner.

As we just discussed, imposing a higher discount rate (i.e., $\delta < 1$) can increase firm value by reducing the overinvestment costs of separation when $\theta = \beta$. However, within this setting the firm fails to achieve V^* (i.e., the firm value when θ is observable) since it leads to underinvestment when $\theta = 1$. We conclude this section by considering whether a policy of multiple hurdle rates (i.e., rates set as a function of the amount of capital invested) will solve the problem. Proposition 10 confirms that this is indeed the case:

Proposition 10 *For each β there exists δ^*_β such that the firm follows the optimal investment policy: $k_1^{1*} = k_1^{s*}$ and $k_\beta^{\delta_\beta^*} = k_\beta^{s*}$ and its value reaches V^*.*

Proposition 10 implies that a policy that imposes higher hurdle rates for larger in-

[29]In separation, a distorted δ may create value by enlarging the set of incentive compatible CEO policies. Since the pooling choices are made by the CEO before observing the type, in pooling distorting δ does not have the potential to increase firm value.

vestments eliminates overinvestment when $\theta = \beta$ without inducing underinvestment when $\theta = 1$. The proposition is thus consistent with the evidence presented in Ross (1986); firms impose hurdle rates which increase with the size of the investment project. Nevertheless other practical considerations outside this model could make a multiple rate policy hard to implement. For instance, in the presence of multiple projects, multiple divisions or projects that require staged investments having a rate which depends on the amount invested could lead the CEO to take actions that "game" the nonlinearity inherent in the multiple hurdle rate policy.

5.2.2 Optimal CEO compensation

From the analysis in the previous section there arises the natural question of whether it is possible to design a compensation contract for the CEO that fully solves the overinvestment problem. In the specific setting that we consider in the paper the answer is yes. As stated in Proposition 10, a compensation policy based on EVA and multiple hurdle rates restores an optimal investment policy.[30]

In practice there are reasons to be skeptical about the ability of CEO compensation to completely solve overinvestment distortions in corporations. On the one hand, to the extent that the board can secretely recontract with the CEO, it is advantageous to the board to secretely induce the CEO to overinvest to induce the agent to exert more effort. For this reason, if unobserved side payments are possible, overinvestment might be required to convey information to the agent (i.e., the separating equilibrium considered in the entrepreneurial model would be restored). On the other hand, even in the absence of hidden compensation, incentive problems at the CEO level may require a pay for performance compensation which will produce effect similar to those explored in previous section, most importantly the need to use the firm investment policy to convey information to the stakeholders.

6 Concluding remarks

This paper proposes a "top-down" theory of capital budgeting where top manager's actions motivate lower managers and the firm's stakeholders to exert effort on the firm's behalf.

[30] More generally, the setting also allows the trivial solution of paying the CEO a fixed amount and inducing truthtelling from the CEO at no cost.

One of the main purposes of our approach is to help reconcile several disconnects between academic theory and industry practice about capital budgeting practices. For example, while the NPV rule proposes using expected cash flow estimates and expected return on investments with equivalent risk to discount those expected cash-flows, in practice there is a tendency to "inflate" cash flow estimates, which are then discounted with "inflated" hurdle rates.

A central premise of our paper is that the communication process between managers, headquarters and the firm's other constituencies has a direct influence on the capital budgeting process because the information conveyed by investment choices can have an important influence on the stakeholders' choices. To explore the interaction between investments expenditures and stakeholder choices, we have examined two related models. First, we considered an entrepreneurial firm whose production process requires a capital expenditure by the firm and effort by the employee. In this setting, we show that there is a natural tendency for firms to overinvestment which in turn can lead to commitments to rigidities in investment policies, like maximum and minimum levels of investment. We then modified the analysis by introducing a third party who either contributes to finance the firm (i.e., debtholders) or has the authority to set specific capital budgeting policies and to offer compensation contracts to the entrepreneur (i.e., board). Within this setting we find that high hurdle rates, the use of debt financing and the use of EVA managerial compensation can help to offset the overinvestment tendencies that emanate from firms with private information about their prospects.

Our theory provides an alternative explanation for the observed overinvestment in firms. While managerial private benefits are likely to provide impetus for excessive investment, in this paper we show that overinvestment can be a second best response of a manager who needs to communicate information to their subordinates or other stakeholders. While our model provides a unique rationale for overinvestment in firms in which agency problems are not a consideration (i.e., firms in which ownership and control are not separated), more generally, the empirical relevance of our theory vis-a-vis a theory based on managerial private benefits is an open question that we leave for future research.

Appendix: Proofs and other technical derivations

Proof of Proposition 1
Let $w_0 = \{w_{1,0}, w_{\beta,0}\}$ be the wages when $z = 0$ and $\theta = \{1, \beta\}$. (In the text, we define $w = \{w_1, w_\beta\}$ when $z = r$ and $\theta = \{1, \beta\}$.) Thus the principal's problem is:

$$\max_{k,w,w_0,e} \quad V = \sum_{\theta=\{1,\beta\}} p_\theta \left[(rk_\theta - w_\theta)\theta e_\theta - w_{\theta,0}(1 - \theta e_\theta) - \frac{1}{2}k_\theta^2 \right] \tag{39}$$

$$\text{s.t.:} \qquad e_\theta = \arg\max_e \{w_\theta \theta e + w_{\theta,0}(1 - \theta e) - \frac{1}{2}k_\theta c e^2\}, \qquad \text{for } \theta = \{1, \beta\} \tag{40}$$

$$w_\theta \theta e_\theta + w_{\theta,0}(1 - \theta e_\theta) - \frac{1}{2}k_\theta c e_\theta^2 \geq 0, \qquad \text{for } \theta = \{1, \beta\}. \tag{41}$$

We prove $w_{\theta,0} = 0$ by contradiction. If $w_{\theta,0} > 0$ and $w_\theta = 0$ then setting $w_{\theta,0} = 0$ increases firm value and induces a higher level of agent's effort. Alternatively, if $w_{\theta,0} > 0$ and $w_\theta > 0$ then reducing both $w_{\theta,0}$ and w_θ while keeping $(w_\theta - w_{\theta,0})$ constant increases firm value without affecting the agent's effort incentives. Finally, we impose $w_{\theta,0} = 0$ and obtain k_θ^*, w_θ^* and e_θ^* by substituting the first order condition of e_θ (40) and then solving in the first order conditions of w_θ and k_θ.

Proof of Proposition 2
If $\beta \geq \beta^*$ then the unconstrained solution described in (7) satisfies $V_\beta^\beta \geq V_\beta^1$ (i.e., IC_β) and $V_1^1 \geq V_1^\beta$ (i.e., IC_1 which can be expressed as $\beta^3(2 - \beta) \geq 1$ or $\beta \geq \beta^*$.) If instead $\beta < \beta^*$ the unconstrained solution does not satisfy IC_1 which requires us to solve the general problem case in which $w_{\theta,0}^s \geq 0$. Let $w_0^s = \{w_{1,0}^s, w_{\beta,0}^s\}$ be the wage when $z = 0$ and $\theta = \{1, \beta\}$. Ignoring (by now) IC_β we get:

$$\max_{w_0^s, w^s, k^s, e^s} \; (1 - \pi)V_1^1 + \pi V_\beta^\beta$$

$$\text{s.t.:} \qquad e_\theta^s = \arg\max_e \{w_\theta^s \theta e + w_{\theta,0}^s(1 - \theta e) - \frac{1}{2}k_\theta^s c e^2\} \qquad \text{for } \theta = \{1, \beta\} \tag{42}$$

$$w_\theta^s \theta e_\theta^s - \frac{1}{2}k_\theta^s c {e_\theta^s}^2 \geq 0 \qquad \text{for } \theta = \{1, \beta\} \tag{43}$$

$$V_1^1 \geq V_1^\beta, \tag{44}$$

where: $V_\theta^{\theta'} \equiv rk_{\theta'}^s \theta e_{\theta'}^s - [w_{\theta',0}^s + (w_{\theta'}^s - w_{\theta',0}^s)\theta e_{\theta'}^s] - \frac{1}{2}{k_{\theta'}^s}^2$.
In the previous problem, type 1's payoff is maximized as in the unconstrained case (i.e., $w_{1,0}^{s*} = 0$, $w_1^{s*} = w_1^*$, and $k_1^{s*} = k_1^*$) because any deviation (i.e., $w_{1,0}^s \neq 0$, $w_1^s \neq \frac{r}{2}k_1^s$, or $k_1^s \neq k_1^*$) reduces $V_1^{1'}$ without easing IC_1.[31] We impose such values and solve for type β's optimal values:

$$\max_{w_{\beta,0}, w_\beta, k_\beta, e_\beta} (1 - \pi)V_1^1 + \pi V_\beta^\beta$$

$$\text{s.t.:} \qquad e_\beta \in \arg\max_e \{w_{\beta,0} + (w_\beta - w_{\beta,0})\beta e - k_\beta c e^2\}, \tag{45}$$

$$V_1^1 \geq V_1^\beta. \tag{46}$$

[31] For ease of notation, we drop the superscript "s".

Expression (46) can be writen as:

$$V_1^1 = \frac{1}{2}k_1^{*2} \geq \{rk_\beta - (w_\beta - w_{\beta,0})\}e_\beta - \frac{1}{2}k_\beta^2 - w_{\beta,0} = V_1^\beta. \tag{47}$$

We define $\alpha_\beta \equiv \frac{w_\beta - w_{\beta,0}}{rk_\beta}$ and express (45) as $e_\beta = \frac{\alpha_\beta r\beta}{c}$ and plug them into objective function:

$$\max_{w_\beta^s,\, k_\beta, \alpha_\beta} (1-\pi)\frac{k_1^{*2}}{2} + \pi\{rk_\beta(1-\alpha_\beta)\beta\frac{\alpha_\beta r\beta}{c} - \frac{1}{2}k_\beta^2 - w_{\beta,0}\},$$

whose Lagrangian is:

$$L = \frac{(1\text{-}\pi)k_1^{*2}}{2} + \pi\{\frac{rk_\beta(1\text{-}\alpha_\beta)\beta^2\alpha_\beta r}{c} - \frac{k_\beta^2}{2} - w_{\beta,0}\} - \lambda\{\frac{rk_\beta(1\text{-}\alpha_\beta)\alpha_\beta r\beta}{c} - \frac{k_\beta^2}{2} - w_{\beta,0} - \frac{k_1^{*2}}{2}\} + \mu w_{\beta,0} + \upsilon k_\beta.$$

By Kuhn-Tucker Theorem, the FOCs with respect to $w_{\beta,0}, \alpha_\beta$, and k_β are:

$$\lambda - \pi + \mu = 0 \tag{48}$$
$$(\pi\beta - \lambda)(1 - 2\alpha_\beta) = 0 \tag{49}$$
$$(\pi\beta - \lambda)r(1-\alpha_\beta)\frac{\alpha_\beta r\beta}{c} - (\pi - \lambda)k_\beta + \upsilon = 0 \tag{50}$$

Since by (48), $\pi\beta - \lambda = \mu + (\beta - 1)\pi > 0$, then (49) implies that $\alpha_\beta = \frac{1}{2}$ and (50) that $(\pi - \lambda)k_\beta > 0$. This means that $\pi - \lambda = \mu > 0$ which implies that $w_{\beta,0} \geq 0$ is binding. Equation (50) also implies that $k_\beta = \frac{(\pi\beta-\lambda)r(1-\alpha_\beta)\frac{\alpha_\beta r\beta}{c}+\upsilon}{\pi-\lambda} > r(1-\alpha_\beta)\frac{\alpha_\beta r\beta}{c} = k_\beta^*$ (i.e., type β overinvests). Solving for k_β in the binding (47) after substituting $\alpha_\beta = \frac{1}{2}$ and $w_{\beta,0} = 0$, we have

$$rk_\beta\frac{r\beta}{4c} - \frac{1}{2}k_\beta^2 - \frac{k_1^{*2}}{2} = 0 \tag{51}$$

which has two real roots. But since $V_\beta^\beta = \beta rk_\beta\frac{r\beta}{4c} - \frac{1}{2}k_\beta^2 = rk_\beta\frac{r\beta}{4c}(\beta - 1) + rk_\beta\frac{r\beta}{4c} - \frac{1}{2}k_\beta^2$, we get:

$$V_\beta^\beta = rk_\beta\frac{r\beta}{4c}(\beta - 1) + \frac{k_1^{*2}}{2} \tag{52}$$

which follows because of (51). Since V_β^β is increasing in k_β, the larger root is optimal i.e., $k_\beta^{s*} = \frac{k_\beta^* + \sqrt{k_\beta^{*2} - k_1^{*2}\beta^2}}{\beta}$. Substituting $k_\theta^* = \frac{r^2\theta^2}{4c}$, we get $k_\beta^{s*} = \frac{1+\sqrt{1-1/\beta^2}}{\beta}k_\beta^*$. Notice that $\frac{1+\sqrt{1-1/\beta^2}}{\beta} = 1$ has two real roots, 1 and β^*. However, since β^* is the largest real root, since $\lim_{\beta\to\infty}\frac{1+\sqrt{1-1/\beta^2}}{\beta} < 1$ and since $\lim_{\beta\to 1}\frac{1+\sqrt{1-1/\beta^2}}{\beta} > 1$, it follows that $\frac{1+\sqrt{1-1/\beta^2}}{\beta} > 1$ for $1 < \beta < \beta^*$. Finally, to check if IC$_\beta$ holds, notice that (52) and $V_1^1 = \frac{k_1^{*2}}{2}$ implies that IC$_\beta$ holds iff $k_\beta^{s*}\beta \geq k_1^*$. This follows since $\beta > 1$ and $k_\beta^{s*} > k_\beta^* > k_1^*$. The principal's payoff (15) follows from plugging in the firm's objective function.

Proof of Proposition 4

While $\beta \geq \beta^*$ separation dominates, both pooling and separation can be optimal when $\beta < \beta^*$. To compare V_s^* with V_p^*, we rewrite $V_s^* = \frac{k_1^{*2}}{2} + \pi(1 - \frac{1}{\beta})\Delta k_\beta^{*2}$. Thus, pooling dominates separation iff: $\frac{\bar{k}^{*2}}{2} - \frac{k_1^{*2}}{2} \geq \pi(1-\frac{1}{\beta})\Delta k_\beta^{*2}$. Substituting $\bar{k}^* = \frac{r^2}{4c}\bar{\theta}^2$, $k_\theta^* = \frac{r^2}{4c}\theta^2$ and $\Delta = \frac{\beta+\sqrt{\beta^2-1}}{\beta^2}$, we get $\frac{[1+\pi(\beta-1)]^4}{2} - \frac{1}{2} \geq \pi(1-\frac{1}{\beta})\beta^2(\beta+\sqrt{\beta^2-1})$ which holds when $\frac{[1+\pi(\beta-1)]^4}{2} - \frac{1}{2} - \pi(\beta-1)[\beta^2 + \beta\sqrt{\beta^2-1}] \equiv \zeta \geq 0$.

To sign ζ, we fix β and define $\pi^*(\beta)$ such that $\zeta = 0$. Notice that $\frac{\partial\zeta}{\partial\pi}\Big|_{\pi=0} = (\beta-1)[2-\beta(\beta+\sqrt{\beta^2-1})]$. Since $[2-\beta(\beta+\sqrt{\beta^2-1})] \geq 0$ iff $\beta \leq \frac{2}{\sqrt{3}}$, then $\frac{\partial\zeta}{\partial\pi}\Big|_{\pi=0} \geq 0$ iff $\beta \leq \frac{2}{\sqrt{3}}$. In this case, since $\zeta(\pi)$ is convex in π, then $\frac{\partial\zeta}{\partial\pi} > 0$. Since $\zeta(\pi=0)=0$, it follows that $\zeta > 0$ for all $\pi > \pi^*(\beta) = 0$ on $(1, \frac{2}{\sqrt{3}}]$.

If $\beta > \frac{2}{\sqrt{3}}$, then $\frac{\partial\zeta}{\partial\pi}\Big|_{\pi=0} < 0$. Also notice that $\zeta(\pi=1) > 0$ because $V_p^*(\pi=1) = \frac{k_\beta^2}{2}$ achieves the full information payoff. Thus by continuity, there exists a $\pi^*(\beta) > 0$ such that $\zeta = 0$. Further, by the convexity of $\zeta(\pi)$, $\zeta(\pi) \leq 0$ for any $0 \leq \pi \leq \pi^*(\beta)$. In addition, because $0 = \zeta(\pi^*) - \zeta(0) = \int_0^{\pi^*} \frac{\partial\zeta}{\partial\pi}(\pi)d\pi < \int_0^{\pi^*} \frac{\partial\zeta}{\partial\pi}(\pi^*)d\pi = \frac{\partial\zeta}{\partial\pi}(\pi^*)\pi^*$,we have $\frac{\partial\zeta}{\partial\pi}\Big|_{\pi=\pi^*(\beta)} > 0$.That is, $\frac{\partial\zeta}{\partial\pi} > 0$ for all $\pi > \pi^*(\beta)$ by convexity and hence $\zeta > 0$ for all $\pi > \pi^*(\beta)$.

Consider next $\zeta(\beta, \pi^*(\beta)) = 0$. After some derivations we can get $\frac{\partial\zeta}{\partial\beta}\Big|_{\pi=\pi^*(\beta)} < 0$ which applying the Implicit Function Theorem to $\zeta(\beta, \pi^*(\beta))$ leads to:

$$\frac{d\pi^*}{d\beta} = -\frac{\frac{\partial\zeta}{\partial\beta}\Big|_{\pi=\pi^*}}{\frac{\partial\zeta}{\partial\pi}\Big|_{\pi=\pi^*}} > 0. \tag{53}$$

Thus, $\pi^*(\beta)$ is strictly increasing on $(\frac{2}{\sqrt{3}}, \beta^*)$.

Let $\beta^{**}(\pi)$: $(0,1) \to (\frac{2}{\sqrt{3}}, \beta^*)$ be the inverse function of $\pi^*(\beta)$. For a fixed π, we can show that $\zeta(\beta) < 0$ iff $\beta^{**}(\pi) < \beta < \beta^*$. To do so, suppose $\beta_1 > \beta^{**}$ and notice that $\zeta(\beta_1, \pi^*(\beta_1)) = 0$. Since $\pi^*(\cdot)$ is strictly increasing, $\pi^*(\beta_1) > \pi$. This implies that for (π, β_1), separation is optimal. The case where $\beta_1 < \beta^{**}$ can be symmetrically proved.

Proof of Lemma 1

First, as in the proof of Proposition 2 we can obtain: (i) $w_\theta^{d*} = \frac{rk_\theta^{d*}-d}{2}$; (ii) wage is zero when the output is zero regardless whether IC_1 is binding or not, (iii) only IC_1 can bind and (iv) $V_{1,d}^1$ reaches the full information payoff. As in the proof of Proposition 1, for type 1: $(w_1^{d*}, k_1^{d*}, e_1^{d*}) = (\frac{rk_1^{d*}-d}{2}, \frac{r^2}{4c} - \frac{d^2}{4ck_1^{d*2}}, \frac{w_1^{d*}}{ck_1^{d*}})$. Since $k_1^{d*} < k_1^{s*}$, it follows that lemma 1 holds for type 1.

Now we show that the IC_1 does not bind and that $k_\beta^d = k_\beta^{s*}$ and k_1^d maximize $V_{1,d}^1$. When $d = 0$, IC_1 is:

$$\left(\frac{rk_\beta^{s*}}{2}\right)^2 \frac{\beta}{ck_\beta^{s*}} - \frac{1}{2}k_\beta^{s*^2} \leq \left(\frac{rk_1^{s*}}{2}\right)^2 \frac{1}{ck_1^{s*}} - \frac{1}{2}k_1^{s*^2} \tag{54}$$

But $\left(\frac{rk_\beta^{s*}\text{-}d}{2}\right)^2 \frac{\beta}{ck_\beta^{s*}} = \left(\frac{rk_\beta^{s*}}{2}\right)^2 \frac{\beta}{ck_\beta^{s*}} - \int_0^d \left(\frac{rk_\beta^{s*}\text{-}x}{2k_\beta^{s*}}\right)\frac{\beta}{c}dx$; $\left(\frac{rk_1^{s*}\text{-}d}{2}\right)^2 \frac{1}{ck_1^{s*}} = \left(\frac{rk_1^{s*}}{2}\right)^2 \frac{1}{ck_1^{s*}} - \int_0^d \left(\frac{rk_1^{s*}\text{-}x}{2k_1^{s*}}\right)\frac{1}{c}dx$ and $\left(\frac{rk_\beta^{s*}\text{-}x}{2k_\beta^{s*}}\right)\frac{\beta}{c} > \left(\frac{rk_1^{s*}\text{-}x}{2k_1^{s*}}\right)\frac{1}{c}$ because $k_\beta^{s*} > k_1^{s*}$. Therefore:

$$\left(\frac{rk_\beta^{s*}-d}{2}\right)^2 \frac{\beta}{ck_\beta^{s*}} - \frac{1}{2}k_\beta^{s*^2} < \left(\frac{rk_1^{s*}-d}{2}\right)^2 \frac{1}{ck_1^{s*}} - \frac{1}{2}k_1^{s*^2} \leq V_{1,d}^1 \tag{55}$$

The second equation follows because k_1^{s*} may not be optimal when $d>0$. Expression (55) is IC$_1$ under $d>0$ after we substituting $w_\theta^{d*}=\frac{rk_\theta^d-d}{2}$. Notice that because k_β^{s*} is (weakly) over-investment for type β when $d=0$, it is always strictly overinvestment for type β when $d>0$. This is because under full information $k_\beta^{d*}=(\frac{r^2}{4c}-\frac{d^2}{4ck_\beta^{d*2}})\beta^2<k_\beta^{s*}$. Since k_β^{s*} satisfies IC$_1$ with strict inequality, $k_\beta^{d*}<k_\beta^{s*}$ reduces overinvestment. Finally $w_\beta^{d*}<w_\beta^{s*}$, $e_\beta^{d*}<e_\beta^{s*}$ follows from $w_\theta^{d*}=\frac{rk_\theta^d-d}{2}$ and $e_\theta^{d*}=\frac{w_\theta^{d*}\theta}{ck_\theta^{d*}}$.

Proof of Proposition 7

Taking derivatives: $\left.\frac{dV^d}{dd}\right|_{d=0}=\left.\frac{\partial V^d}{\partial k_\beta^d}\right|_{d=0}\left.\frac{dk_\beta^d}{dd}\right|_{d=0}$. Since by assumption $\left.\frac{\partial V^d}{\partial k_\beta^d}\right|_{d=0}<0$ and by Lemma 1 $\left.\frac{dk_\beta^d}{dd}\right|_{d=0}<0$ then $\left.\frac{dV^d}{dd}\right|_{d=0}>0$. The second part of the proposition is immediate.

Proof of Proposition 9

We first show that $\delta=1$ is not optimal. If we derive V_δ with respect to δ we get: $\frac{dV_\delta}{d\delta}=\frac{\partial V_\delta}{\partial\alpha_\delta(1)}\frac{d\alpha_\delta(1)}{d\delta}+\frac{\partial V_\delta}{\partial\alpha_\delta(\beta)}\frac{d\alpha_\delta(\beta)}{d\delta}+\frac{\partial V_\delta}{\partial k_\delta(1)}\frac{dk_\delta(1)}{d\delta}+\frac{\partial V_\delta}{\partial k_\beta^\delta}\frac{dk_\delta(\beta)}{d\delta}$.
Since $\alpha_\delta(\theta)$ satisfies the FOC $\left.\frac{\partial V_s}{\partial\alpha_\theta}\right|_{\alpha(\theta)=\frac{1}{2}}=\left.\frac{\partial V_\delta}{\partial\alpha_\theta^\delta}\right|_{\delta=1}=0$. Similarly, $\frac{\partial V_\delta}{\partial k_1^\delta}=0$ at $\delta=1$, since k_1^{s*} satisfies FOC $\left.\frac{\partial V_s}{\partial k_1^s}\right|_{k_1^s=k_1^{s*}}=\left.\frac{\partial V_\delta}{\partial k_1^\delta}\right|_{\delta=1}=0$. Finally, $\left.\frac{\partial V_s}{\partial k_\beta^s}\right|_{k_\beta^s=k_\beta^{s*}}=\left.\frac{\partial V_\delta}{\partial k_\beta^\delta}\right|_{\delta=1}<0$ because $k_\beta^{s*}>k_\beta^*$. Therefore, since $\frac{dk_\delta(\beta)}{d\delta}>0$, then $\left.\frac{dV_\delta}{d\delta}\right|_{\delta=1}=\left.\frac{\partial V_\delta}{\partial k_\beta^\delta}\right|_{\delta=1}\left.\frac{dk_\delta(\beta)}{d\delta}\right|_{\delta=1}<0$.
We then show that $\delta>1$ cannot be optimal. First we check that the optimal contract under $\delta>1$, $(w_\theta^\delta,k_\theta^\delta)$ satisfies IC$_1$ under $\delta=1$. When $\delta>1$, IC$_1$ is

$$\delta^4k_\beta^{s*}r(1-\frac{1}{2})e_\beta^{s*}-\delta^4\frac{k_\beta^{s*2}}{2}\leq\delta^4k_1^{s*}r(1-\frac{1}{2})e_1^{s*}-\delta^4\frac{k_1^{s*2}}{2}. \tag{56}$$

Plugging in $(w_\theta^\delta,k_\theta^\delta)$ when $\delta=1$, the IC$_1$ becomes

$$\delta^3k_\beta^{s*}r(1-\frac{\delta}{2})e_\beta^{s*}-\delta^4\frac{k_\beta^{s*2}}{2}\leq\delta^3k_1^{s*}r(1-\frac{\delta}{2})e_1^{s*}-\delta^4\frac{k_1^{s*2}}{2}, \tag{57}$$

Comparing both inequalities, we find that deducting $\delta^3k_\beta^{s*}r(\delta-1)e_\beta^{s*}$ from the LHS and deducting $\delta^3k_1^{s*}r(\delta-1)e_1^{s*}$ on the RHS of (56) yields (57). But

$$\delta^3k_\beta^{s*}r(\delta-1)e_\beta^{s*}>\delta^3k_1^{s*}r(\delta-1)e_1^{s*}, \tag{58}$$

iff $\delta>1$. Therefore, (57) is satisfied with strict equality. We can also check that $(w_\theta^\delta,k_\theta^\delta)$ satisfies all the constraints in program (45) to (46). As a result, $(w_\theta^\delta,k_\theta^\delta)$ is strictly suboptimal under $\delta=1$, since at the optimum IC$_1$ must bind. Hence, firm value when under $(w_\theta^\delta,k_\theta^\delta)$ is less than $V_{\delta=1}=V_s^*$. Technically, reducing δ from 1 relaxes IC$_1$ constraint, which is binding at $\delta=1$. This can be seen by noticing that the inequality in (58) flips for $\delta<1$. Therefore, (57) implies (56).

Technical derivation: The optimal mechanism

Before proceeding to the optimal contract design, we need point out that the Revelation Principle in the mechanism design literature (Laffont and Green (1977), Myerson (1979), and Dasgupta, Hammand, and Maskin (1979) fails to hold in this setting as shown by Proposition 4. This is because the agent's effort choice is a function of the principal's announcement. That is, the agent's effort choice has to maximize the agent's payoff conditional on the principal's announcement. This incentive compatibility constraint breaks the equivalence between the direct mechanism, in which the principal tells the truth, and the indirect mechanism in which the principal does not.turns out that the optimal mechanism can be achieved by either pooling or separation. To show that, we first show that it is without loss of generality to consider only two contracts in our mechanism design. Denote a contract $c_i \equiv (w_i(q), k_i)$. Since revelation principle does not hold, choosing a contract from the menu $\{c_i\}_{i\in I}$ may not fully reveal his type.[32] Let $\psi(\theta)$ be the set of contracts that θ chooses with positive probability (and for a slight abuse of notation, it also includes the probability of each type choosing c_i, $\pi_\theta(c_i)$ for $c_i \in \psi(\theta)$). We denote $w_{i,0} = w_i(0)$ and $w_i = w_i(rk_i)$ Formally, the principal's problem is

$$\max_{c_i, i\in I} \sum_{\theta\in\{1,\beta\}} \Pr(\theta) \sum_{c_i\in\psi(\theta)} \pi_\theta(c_i)V_\theta(c_i)$$

$$\text{s.t.:} \quad e_i \in \arg\max_e \{w_{i,0} + (w_i - w_{i,0})E[\theta|c_i] - h(e, k_i)\}, \tag{59}$$

$$w_{i,0} + (w_i - w_{i,0})E[\theta|c_i] - h(e, k_i) \geq 0, \tag{60}$$

$$V_\theta(c_i) \geq V_\theta(c_j), \text{ for any } c_i \in \psi(\theta) \tag{61}$$

$$V_\theta(c_i) \geq 0 \tag{62}$$

$$\sum_i \pi_\theta(c_i) = 1, \tag{63}$$

$$w_i \geq 0, w_{i,0} \geq 0 \tag{64}$$

Let $\alpha_i = \frac{w_i - w_{i,0}}{k_i r}$ and $a_i \equiv rk_i(1-\alpha_i)e_i$. By the standard revealed preference argument, we get:

Lemma 3 (Monotonicity) In any mechanism: (i) If $\theta = \beta$ picks c_i with positive prob. and $\theta = 1$ picks c_j with positive prob. then $a_i \geq a_j$; and (ii) If $\theta = \beta$ (respectively $\theta = 1$) is indifferent between c_i and c_j with $a_i > a_j$, then $\theta = 1$ (respectively $\theta = \beta$) strictly prefers c_j to c_i.

Proof. Type 1 and β's ICs imply: $a_i\beta - \frac{1}{2}k_i^2 - w_{i,0} \geq a_j\beta - \frac{1}{2}k_j^2 - w_{j,0}$ and $a_j - \frac{1}{2}k_j^2 - w_{j,0} \geq a_i - \frac{1}{2}k_i^2 -^0 w_{i,0}$. Adding up both inequalities yields

$$(a_i - a_j)(\beta - 1) \geq 0, \tag{65}$$

which implies (i). For part (ii) we first consider type $\beta^{'}$s IC . Indifference implies $\beta(a_i - a_j) = \frac{1}{2}k_i^2 - \frac{1}{2}k_j^2 + w_{i,0} - w_{j,0}$. That is $(a_i - a_j) < \frac{1}{2}k_i^2 - \frac{1}{2}k_j^2 + w_{i,0} - w_{j,0}$, which implies that type 1 strictly prefers c_j. The case where type 1 is indifferent follows by similar argument.

[32] I need not be finite or even countable.

Lemma 4 For any mechanism, there is another mechanism with at most two contracts that is no worse for each type..

Proof. There are two cases. First, in the original mechanism there are two contracts $c_\theta \in \{1, \beta\}$ such that c_θ is chosen by type θ only. In this case, a separation mechanism with only two contracts, $c_\theta \in \{1, \beta\}$ and type θ only chooses c_θ with probability 1 will be no worse than the original mechanism. The reason is that for each type, $V_\theta(c_\theta)$ is the best value it can achieve in the original mechanism by its IC and in the new mechanism it achieves it with probability 1. Notice that c_θ satisfies all the constraints from (59) to (64).

Second case: one type always pools with the other. We first consider the case where this pooling type is β. More specifically, type 1 chooses any contract in $\psi(\beta)$ with positive probability. The probability that a type 1 chooses any contract from $\psi(\beta)$ is $\pi_1(\psi(\beta)) \leq 1$. By lemma 3, if there are two contracts c_i and c_j in $\psi(\beta)$, then $a_i = a_j = a$. We now construct the new mechanism. We first construct a contract c'_p for both types to pool. Pick a contract, c_p, in $\psi(\beta)$ such that $E[\theta|c_p] \leq E[\theta|\psi(\beta)]$, where $E[\theta|\psi(\beta)] \equiv E[E[\theta|c_i]|c_i \in \psi(\beta)]$, is the conditional expectation of θ on a contract $c_i \in \psi(\beta)$ is chosen. We then modify c_p to c'_p. We let type β choose c'_p with probability 1 and type 1 choose it with probability $\pi_1(\psi(\beta))$. Thus, $E[\theta|c'_p] = E[\theta|\psi(\beta)]$. We will modify k'_p so that $a'_p = k'_p r(1-\alpha_p)\frac{\alpha_p E[\theta|c'_p]}{c} = k_p r(1-\alpha_p)\frac{\alpha_p E[\theta|c_p]}{c} = a$. Since $E[\theta|c'_p] \geq E[\theta|c_p]$, $k'_p \leq k_p$. We then increase $w_{p,0}$ to $w'_{p,0}$ so that $\frac{1}{2}k_p^2 + w_{p,0} = \frac{1}{2}k_p'^2 + w'_{p,0}$. Thus $c'_p = \{k'_p, \alpha_p, w'_{p,0}\}$. We can check that both type's payoff from choosing c'_p are the same as that of choosing a contract in $\psi(\beta)$ in the original mechanism. Now we look at type 1. There are two cases. If $\pi_1(\psi(\beta)) = 1$, then we are done because the original mechanism is the same as this one contract, c'_p with both type pooling on it. If $\pi_1(\psi(\beta)) < 1$, then exists a contract c_1 so that only type 1 chooses with positive probability in the original mechanism. We now keep c_1 in the new mechanism and let type 1 chooses it with probability $1 - \pi_1(\psi(\beta))$. The new mechanism has tow contracts, c'_p and c_1. We can see that type 1 is indifferent between the two and type β strictly prefers c'_p. Each type gets the same payoff as in the original mechanism. The case where the pooling type is 1 is proved similarly.

Proposition 11 The principal's maximum expected payoff is achieved with either pooling or separation.

Proof. We only need to consider $\beta < \beta^*$. For $\beta \geq \beta^*$, separation achieves the full information payoff. By lemma 4, we only need to consider two contracts as in the separation case. However, each type of principal may select the contracts randomly, i.e., revealing information partially. Denote c_θ the cotract a type θ chooese with positive probability. Now the mechanism is $(c_{\widehat{\theta}}, \sigma_\theta)$, As before, $\hat{\theta}$ is the principal's claim and θ is her true type. σ_θ is the probability that a type θ principal chooses c_θ. The value of a type θ from choosing $c_{\widehat{\theta}}$ is $V_\theta^{\hat{\theta}}$. Formally the contract design problem is

$$\max_{w_{\widehat{\theta},0}, w_{\widehat{\theta}},\, k_{\widehat{\theta}},\, \sigma_\theta} \sum_{\theta=1,\beta} p_\theta V_\theta^\theta$$

$$\text{s.t.:} \qquad e_{\hat{\theta}} \in \arg\max_e \{w_{\widehat{\theta},0} + (w_{\hat{\theta}} - w_{\hat{\theta},0}) E[\theta|\widehat{\theta}]e - k_{\hat{\theta}} f(e)\}, \tag{66}$$

$$w_{\hat{\theta},0} + (w_{\hat{\theta}} - w_{\hat{\theta},0}) E[\theta|\widehat{\theta}] e_{\hat{\theta}} - k_{\hat{\theta}} f(e_{\hat{\theta}}) \geq 0, \tag{67}$$

$$V_\theta^\theta \geq V_\theta^{\hat{\theta}}, \text{ for any } \hat{\theta} \neq \theta \tag{68}$$

$$(1-\sigma_\theta)[V_\theta^\theta - V_\theta^{\hat{\theta}}] = 0, \text{ for } \theta = 1, \beta \tag{69}$$

This problem is similar to optimization problem in the separation case with two differences. First, the mechanism can be random. Second, there is a new constraint, (69), which is the complementary slackness constraint. It says that if type θ is strictly better off choosing a contract, she will choose it with probability 1. On the other hand, if she is indifferent between the two contracts, she can randomize. (69) is a natural requirement for mixed strategies.

Notice that this program includes the full separation and full pooling equilibria studied before. This is because for the separation case, $\sigma_\theta = 1$. In the pooling case, c_θ is the same for all θ. In general, we need to consider three cases: I. Only type β's IC of (68) is binding; II. only type 1's IC of (68) is binding; and III. both types ICs are binding.

Case I is impossible for the following reason. In this case, we can solve the problem with only type β's IC (ignoring type 1's IC and the associated complementary slackness constraint). However, we know that the full information allocation in subsection (3.1) satisfies all the remaining constraints. So the optimal solution should be the full information allocation (it should be clear that the program cannot get a better solution than the full information case). We know that is not possible since the full information allocation violates type 1's IC by assumption $\beta < \beta^*$.

Case III cannot be better than the pooling equilibrium studied in subsection 3.4. In case III, both types of principal are indifferent between the two contracts. Therefore the objective function becomes

$$\max_{w_{1,0}, w_1,\, k_1} \sum_{\theta=1,\beta} p_\theta V_\theta^1$$

that is, the objective function only depends on contract for type 1. Without loss of generality, we assume $E[\theta|c_1] \leq E[\theta]$. That is, if the type 1 contract is chosen, the agent's posterior expectation of θ is lowered (this assumption is without loss of generality because we can always define type 1 contract such that this assumption is satisfied). Now we claim that the solution can never be strictly better than the pooling equilibrium studied in subsection 3.4. Otherwise, let the contract $(w_{1,0}, w_1, k_1)$, be the contract in the pooling equilibrium. The agent would exert a (weakly) higher effort in the pooling case than here when $(w_{1,0}, w_1, k_1)$ is chosen since $E[\theta|c_1] \leq E[\theta]$. The contract satisfies the agent's IR in the pooling case too because (67) is satisfied and $E[\theta|c_1] \leq E[\theta]$. Therefore, under c_1, the principal has a better payoff than under $(\bar{w}^*, \bar{k}^*)$, which is a contradiction to the optimally of the contract $(\bar{w}^*, \bar{k}^*)$.

Therefore, we only need to consider case II as the potential optimal mechanism. In this case, the type 1 principal using a mixed strategy choosing the two contracts, and the type β always chooses $(w_{\beta,0}, w_\beta, k_\beta)$. Because when type 1 chooses c_1 she is fully revealed, the allocation in this case should the same as in the full information case as argued in the proof of proposition 2. Formally, the program is

$$\max_{w_{\beta,0},w_\beta,k_\beta,\sigma_1} (1-\pi)V_1^1 + \pi V_\beta^\beta \tag{70}$$

$$\text{s.t.} \quad e_\beta \in \arg\max_e \{w_{\beta,0} + [w_\beta - w_{\beta,0}]E[\theta|\widehat{\theta}=\beta]e - k_\beta f(e)\}, \tag{71}$$

$$V_1^1 \geq V_1^\beta, \tag{72}$$

$$w_\beta \geq 0, w_{\beta,0} \geq 0 \tag{73}$$

$$k_\beta \geq 0, \tag{74}$$

$$0 \leq \sigma_1 \leq 1. \tag{75}$$

By Bayes rule, $E[\theta|\widehat{\theta}=\beta] = 1 + \frac{\pi}{\pi+(1-\pi)(1-\sigma_1)}(\beta-1)$. Let $E[\theta|\widehat{\theta}=\beta] = v\beta$, $v \in [(\beta)^{-1}, 1]$. Because v and σ_1 are one to one, we can change the control variable σ_1 to v. Consider the Lagrangian of the previous problem:

$$L \equiv \frac{(1\text{-}\pi)k_1^{*2}}{2} + \pi\left(\frac{rk_\beta(1\text{-}\alpha_\beta)\alpha_\beta rv\beta^2}{c} \text{-} g(k_\beta)\text{-}w_{\beta,0}\right) \text{-} \lambda\left(\frac{rk_\beta(1\text{-}\alpha_\beta)\alpha_\beta rv\beta}{c} \text{-} g(k_\beta)\text{-}w_{\beta,0}\text{-}\frac{k_1^{*2}}{2}\right) + \mu w_{\beta,0} + \upsilon k_\beta \text{-} \xi(v\text{-}1)$$

By Kuhn-Tucker Theorem, we get the following FOCs with respect to $w_{\beta,0}$, α_β, k_β, and v, respectively

$$\lambda - \pi + \mu = 0, \tag{76}$$

$$(\pi\beta - \lambda)(1-2\alpha_\beta) = 0, \tag{77}$$

$$(\pi\beta - \lambda)r(1-\alpha_\beta)\frac{\alpha_\beta rv}{c} - (\pi - \lambda)k_\beta + \upsilon = 0, \tag{78}$$

$$(\pi\beta - \lambda)rk_\beta(1-\alpha_\beta)\frac{\alpha_\beta r\beta}{c} - \xi = 0. \tag{79}$$

Similar argument as in the proof of Proposition 2 imply that $\pi\beta - \lambda > 0$ which it turns implies that $\xi > 0$, and $v = 1$ and implies $\sigma_1 = 1$. That is, the optimal solution is actually the separation case as studied before. Thus, it is without loss of generality to focus on only pooling and separating equilibria.

Proofs of Section 4

We consider the general wage in the proofs, i.e., $w(\theta, q)$. We also use $V(\theta)$ to denote the firm's equilibrium value. We let $V(\theta, \widehat{\theta}) = V_{\theta}^{\widehat{\theta}}$, and $k(\theta) = k_{\theta}$. It is useful to define $\alpha(\widehat{\theta}) = \frac{w(\widehat{\theta}, rk(\widehat{\theta})) - w(\widehat{\theta}, 0)}{rk(\widehat{\theta})}$ for a contact c.

Proof of Proposition 5

For ease of notation, we denote $V(\theta) = V_{\theta}^{s}$, $V(\theta, \widehat{\theta}) = V_{\theta}^{\widehat{\theta}}$, and $k(\theta) = k_{\theta}^{s}$. When there is little confusion, we will drop the superscript s in the proof. Proposition 5 follows directly from the following lemma.

Lemma 5 The conditions in Proposition 5 define the separating equilibrium with a highest the equilibrium payoff to the principal, $V(\theta)$.

Proof. In a separating equilibrium, the principal reveals her information, i.e., $\widehat{\theta} = \theta$

$$\theta = \arg\max_{\widehat{\theta}} V(\theta, \widehat{\theta}) = r^2 k(\widehat{\theta})(1 - \alpha(\widehat{\theta}))\alpha(\widehat{\theta})\frac{\widehat{\theta}}{c}\theta - \frac{1}{2}k(\widehat{\theta})^2 \tag{80}$$

where $\alpha(\widehat{\theta}) = \frac{w(\widehat{\theta}, rk(\widehat{\theta})) - w(\widehat{\theta}, 0)}{rk(\widehat{\theta})}$.Notice that FOC of (80) implies that

$$\frac{r^2}{4c}\theta^2 k(\theta) + \frac{r^2}{4c}k(\theta)\theta - k(\theta)\dot{k}(\theta) = 0, \tag{81}$$

$\dot{k}(\theta) = \frac{dk(\theta)}{d\theta}$, with initial value $k(1) = \frac{r^2}{4c}$. For simplicity we solve $\theta(k)$ rather than $k(\theta)$.

$$\theta = \sqrt{\frac{4c(2k(\theta)^3 + k_1^{*3})}{3k(\theta)r^2}}, \tag{82}$$

is the unique positive solution of $\theta(k)$. The SOC of (80) requires that $\left.\frac{\partial V(\theta, \widehat{\theta})}{\partial \widehat{\theta}^2}\right|_{\widehat{\theta}=\theta} \leq 0$. For ease of notation, denote $V_i(\theta, \widehat{\theta})$, $i = 1, 2$, the partial derivative with respect to the first and second argument. $V_{ij}(\theta, \widehat{\theta})$ is the second order partial derivative. FOC implies $V_2(\theta, \theta) = 0$ for all θ. Therefore, $0 = \frac{dV_2(\theta, \theta)}{d\theta} = V_{12}(\theta, \theta) + V_{22}(\theta, \theta)$. Thus: $V_{22}(\theta, \theta) = -V_{12}(\theta, \theta) = -\frac{r^2}{4c}\frac{dk(\theta)\theta}{d\theta} \leq 0$ if $k(\theta)\theta$ is non-decreasing. Furthermore, we can show that if $k(\theta)\theta$ is non-decreasing, it is sufficient for for all θ to tell the truth. Suppose, $V(\theta, \widehat{\theta}) > V(\theta, \theta)$ for $\theta \neq \widehat{\theta}$ or $\int_{\theta}^{\widehat{\theta}} V_2(\theta, x)dx > 0$. Using FOC, this is equivalent to $\int_{\theta}^{\widehat{\theta}} [V_2(\theta, x) - V_2(x, x)]dx > 0$ or $\int_{\theta}^{\widehat{\theta}} \int_{x}^{\theta} V_{12}(y, x)dydx > 0$. If $\widehat{\theta} > \theta$, $x > \theta$, $V_{12}(y, x) = -\frac{r^2}{4c}\frac{dk(x)x}{dx} \leq 0$, the above inequality is impossible to hold. If $\widehat{\theta} < \theta$, $x < \theta$, it is also impossible to hold.

By (81), $\frac{dk(x)x}{dx} = \frac{k(\theta)\dot{k}(\theta)}{\frac{r^2}{4c}\theta}$. So $\frac{dk(x)x}{dx} \geq 0$ if and only if $\dot{k}(\theta) \geq 0$. But $\dot{k}(\theta) = \frac{-\frac{r^2}{4c}k(\theta)\theta}{\frac{r^2}{4c}\theta^2 - k(\theta)}$ by (81),

therefore $\dot{k}(\theta) \geq 0$ if and only if $\frac{r^2}{4c}\theta^2 \leq k(\theta)$. That is, in equilibrium, we must have overinvestment. We only need to show that (82) has an unique root which is bigger than $k^*_\theta = \frac{r^2}{4c}\theta^2$. After some algebra, (82) is equivalent to: $2k^3 + k_1^{*3} - 3k^2 k^*_\theta = 0$with $k = k(\theta)$. Differentiate the LHS w.r.t. k, we get $6k^2 - 6kk^*_\theta$, which is positive if $k > k^*_\theta$. At $k = k^*_\theta$, the LHS is $k_1^{*3} - k_\theta^{*3} \leq 0$. Therefore, we must have a root $k \geq k^*_\theta$. Because the LHS is increasing for $k \geq k^*_\theta$, we can only have one root $k \geq k^*_\theta$. The equality only holds when $\theta = 1$.
To show that this equilibrium gives the highest payoff to the principal we proceed by contradiction. Suppose there is a separating equilibrium such that $V'(\theta_0) > V^s(\theta_0)$ for some θ_0. Because $V^s(\underline{\theta})$ is the second best, we have $V'(\underline{\theta}) \leq V^s(\underline{\theta})$. Hence we have $\theta_0 > \underline{\theta}$. Let $\theta_e = \sup\{\theta | V'(\theta) \leq V^s(\theta)\}$ and $\theta \in [\underline{\theta}, \theta_0)\}$. Applying Mean Value Theorem to $V'(\theta) - V^s(\theta)$ on $[\theta_e, \theta_0]$, we have a $\theta_1 \in [\theta_e, \theta_0]$ such that $\left.\frac{dV'(\theta)}{d\theta}\right|_{\theta_1} > \left.\frac{dV^s(\theta)}{d\theta}\right|_{\theta_1}$. $V'(\theta_1) \geq V^s(\theta_1)$ by definition of θ_1. It is

$$\begin{aligned} r^2 k^s(\theta_1)(1-\alpha^s(\theta_1))\alpha^s(\theta_1)\frac{\theta_1^2}{c} - \frac{1}{2}k^s(\theta_1)^2 &\leq r^2 k'(\theta_1)(1-\alpha'(\theta_1))\alpha'(\theta_1)\frac{\theta_1^2}{c} - \frac{1}{2}k'(\theta_1)^2 - w'(\theta_1, 0) \\ &\leq r^2 k'(\theta_1)(1-\alpha'(\theta_1))\alpha'(\theta_1)\frac{\theta_1^2}{c} - \frac{1}{2}k'(\theta_1)^2 \\ &\leq r^2 k'(\theta_1)(1-\alpha^s(\theta_1))\alpha^s(\theta_1)\frac{\theta_1^2}{c} - \frac{1}{2}k'(\theta_1)^2. \end{aligned} \tag{83}$$

The last inequality follows because $\alpha^s(\theta_1) = \arg\max(1-\alpha)\alpha$. Because $k^s(\theta_1) \geq k^*(\theta_1)$, and $\frac{\partial}{\partial k}[r^2 k(1-\alpha^s(\theta_1))\alpha^s(\theta_1)\frac{\theta_1^2}{c} - \frac{1}{2}k^2] \leq 0$ for all $k \geq k^*(\theta)$, (83) implies that $k^s(\theta) \geq k'(\theta)$. Now compare $\frac{dV(\theta)}{d\theta}$ at $\theta = \theta_1$. In any seperating equilibrium, Envelope Theorem implies that

$$\frac{dV(\theta)}{d\theta} = \frac{dV(\theta,\theta)}{d\theta} = r^2 k(\theta)(1-\alpha(\theta))\alpha(\theta)\frac{\theta}{c}. \tag{84}$$

Therefore, $\left.\frac{dV'(\theta)}{d\theta}\right|_{\theta_1} = r^2 k'(\theta_1)(1-\alpha'(\theta_1))\alpha'(\theta_1)\frac{\theta_1}{c} \leq r^2 k'(\theta_1)(1-\alpha^s(\theta_1))\alpha^s(\theta_1)\frac{\theta_1}{c} \leq r^2 k^s(\theta_1)(1-\alpha^s(\theta_1))\alpha^s(\theta_1)\frac{\theta_1}{c} = \left.\frac{dV^s(\theta)}{d\theta}\right|_{\theta_1}$, which is contradiction.

Proof of Proposition 6

It is helpful to first show that the optimal mechanism is a partition on the type space. On each partition there can be either pooling or separation. Specifically, let $\Psi = \{\varphi_1, \varphi_2, ...\}$ be a partition of the interval $[1, \bar{\beta}]$. That is, $[1, \bar{\beta}] = \underset{i\in\mathbb{N}}{\cup}\varphi_i$, $\varphi_i \cap \varphi_j = \varnothing$, and. $\varphi_i = (\underline{\theta}_i, \overline{\theta}_i]$ (it can be closed or open intervals since it shouldn't affect the result). Notice that the number of subintervals φ_i can be countably infinite. Let $\varphi(\theta) \equiv \varphi_i$ such that $\theta \in \varphi_i$. We first have a technical lemma.

Lemma 6 It is without loss of generality to consider the above partition mechanism.

Proof. For a fixed $a = [rk - (w(kr) - w(0))]e$, denote the set, $\Theta(a)$, of types that choose it with positive probability. If there is only one types in $\Theta(a)$. Then this contract is separation. If there are multiple types, we know from the Monotonicity Lemma that only the boundary points $(\sup\Theta(a)$

and $\inf \Theta(a)$) can choose more than one contract and all types in between have to choose a. Thus we have pooling. Notice that pooling here only means the types choose the same a, it does not imply that the contract are the same. But we can show that it is optimal for the types that choose the same a pool together.[33] So the pooling types form non-overlapping intervals (including, open, close, half open half close) with probability 1. Now we show that separation types also forms union of intervals. Let $\varphi^s_{\theta_i}$ be an interval between two separation intervals with $\theta_i \in \varphi^s_{\theta_i}$.[34] Therefore the separation types are countable disjoint union of $\varphi^s_{\theta_i}$. If such $\varphi^s_{\theta_i}$ does not exists, then it follows that the set of separation is of probability zero.[35] Because we can ignore sets of probability zero types in our mechanism design, we can thus focus on pooling partition only in this case.
At the boundry of a interval, we should have

$$V(\overline{\theta}_i) = V(\underline{\theta}_{i+1}) \tag{85}$$

in order for the mechanism to be incentive compatible. The next result says that we only need to worry about low type mimic high type.

Lemma 7 The maximization problem yields the same solution if we replace (85) with $V(\overline{\theta}_i) \geq V(\underline{\theta}_{i+1})$. In the optimal mechanism, $\alpha(\theta) = \frac{1}{2}$ and $w(\theta, 0) = 0$.

Proof. We first prove the first part. The proof is by contradiction. We first assume that the solution to the modified problem is strictly better than the original problem. We will then construct another solution to the modified problem such that it is better than that of the modified problem.
Suppose $k'(\cdot), \Psi'$, $t'(\cdot)$ are solution to the modified problem that is strictly better than the original solution (since the modified problem is with less constraint, it cannot yield worse solution). Let φ'_i be an interval such that $V(\overline{\theta}_i) > V(\underline{\theta}_{i+1})$. Now we improve the solution for the modified problem. We will show that all the types in φ'_{i+1} are better off if we modify the solution by increasing $V(\underline{\theta}_{i+1})$. This is a contradiction becuase i) we improve the expected payoff of the firm and ii) $V(\overline{\theta}_i) \geq V(\underline{\theta}_{i+1})$ for all i.
First suppose φ'_{i+1} is separation. In this case, we just let $V(\underline{\theta}_{i+1}) = V(\overline{\theta}_i)$, $\alpha(\theta) = \frac{1}{2}$, and $w(\theta, 0) = 0$. We thus can show as in the proof of Proposition 5 that the new separation is better than the original separation in that all types in φ'_{i+1} are better off.
Second, consider φ'_{i+1} is pooling. Let α'_{i+1}, k'_{i+1} and $w'_{i+1}(0)$ be the contract for the types in φ'_{i+1}. We first let α'_{i+1} goes to $\frac{1}{2}$ if $\alpha'_{i+1} \neq \frac{1}{2}$ so as to increase $(1-\alpha'_{i+1})\alpha'_{i+1}$ and thus $V(\theta)$ on φ'_{i+1}

[33] We have shown $a = rk_c(1-\alpha_c)\alpha_c E[\theta|c(\theta)]$ where $c(\theta)$ is the contract chosen by θ. $\alpha_c = \frac{w_c(k_c r) - w_c(0)}{rk_c}$. Therefore, if there are multiple contracts with the same a, the one with least $\frac{1}{2}k_c^2 + w_c(0)$ is optimal for all the types. This implies that for all the contracts with the same a, the $\frac{1}{2}k_c^2 + w_c(0)$ is the same. Futhermore, we can find an c_p such that $a = (1-\alpha_p)\alpha_p E[\theta|a]$. That is, the pooling of all types choosing the same c_p can implement a. The way to find c_p is to pick a contract with the same a such that $E[\theta|c] \leq E[\theta|a]$ (it always exists). If all types choosing a pool at the contract c, then $a' = rk(1-\alpha_c)\alpha_c E[\theta|a)] \geq rk(1-\alpha_c)\alpha_c E[\theta|c] = a$. But we can change α_c to α_p to lower a' to a. Therefore it is without loss of generality to consider one contract for the same a.

[34] More precisely, for a given pooling interval $(\underline{\theta}_i, \overline{\theta}_i]$, we can find a separation $\varphi^s_i = (\sup\{\theta | \theta < \underline{\theta}_i$ and θ is pooling$\}, \underline{\theta}_i]$.

[35] In this case, any open interval cover one end point of a pooling interval would intersect with another pooling intervals. The whole separation is thus covered by the union of such open intervals. Since there are at most countably many such open intervals, the total measure of separation is thus zero.

and we are done. Similar argument show that $w'_{i+1}(0)=0$. If $\alpha'_{i+1}=\frac{1}{2}$ and $w'_{i+1}(0)=0$, we then change k'_{i+1} to increase $V(\underline{\theta}_{i+1})$. There are two ways. If increasing k'_{i+1} can increase $V(\underline{\theta}_{i+1})$, we are done increasing all $V(\theta)$ on φ'_{i+1} because if k'_{i+1} is under-investing for $\underline{\theta}_{i+1}$, it is under-investing for all $\theta>\underline{\theta}_{i+1}$. The other is decreasing k'_{i+1}. In this case, we need to create two pooling partitions on φ'_{i+1}, i.e., $\varphi_{i+1}=[\underline{\theta}_{i+1},\theta_{i+1}]$ and $\varphi_{i+1'}=(\theta_{i+1},\overline{\theta}_{i+1}]$. θ_{i+1} and the investment on φ_{i+1} k_{i+1} are chosen so that

$$\begin{aligned} V(\underline{\theta}_{i+1}|\varphi_{i+1}) &= V(\overline{\theta}_i), \\ V(\theta_{i+1}|\varphi_{i+1}) &= V(\theta_{i+1}|\varphi'_{i+1}). \end{aligned} \tag{86}$$

But can both equations hold? Notice that for any θ_{i+1},we can find k_{i+1} so that (86) holds (more specifically, we have $\max\limits_{k_{i+1}}\frac{rk_{i+1}}{4c}\underline{\theta}_{i+1}E[\theta|\theta\in[\underline{\theta}_{i+1},\theta_{i+1}]]-\frac{1}{2}k_{i+1}^2>\frac{rk_i}{4c}\overline{\theta}_iE[\theta|c(\overline{\theta}_i)]-\frac{1}{2}k_i^2$ because $\underline{\theta}_{i+1}E[\theta|\theta\in[\underline{\theta}_{i+1},\theta_{i+1}]]>\overline{\theta}_iE[\theta|c(\overline{\theta}_i)]$. $c(\overline{\theta}_i)$ is the contract chosen by $\overline{\theta}_i$. So we can increase k_{i+1} till the equation holds). Notice that $k_{i+1}<k'_{i+1}$ because $\underline{\theta}_{i+1}E[\theta|\theta\in[\underline{\theta}_{i+1},\theta_{i+1}]]<\underline{\theta}_{i+1}E[\theta|\theta\in\varphi'_{i+1}]$. If θ_{i+1} is close to $\underline{\theta}_{i+1}$, we will have $V(\theta_{i+1}|\varphi_{i+1})$ close to $V(\underline{\theta}_{i+1}|\varphi_{i+1})$ and $V(\theta_{i+1}|\varphi_{i+1'})$ close to $V(\theta_{i+1}|\varphi'_{i+1})$ by continuity in θ_{i+1}. Since $V(\underline{\theta}_{i+1}|\varphi'_{i+1})<V(\overline{\theta}_i)$ by assumption and (86), we have $V(\theta_{i+1}|\varphi_{i+1})>V(\theta_{i+1}|\varphi'_{i+1})$. Therefore, if we cannot have both equations hold, we will have (86) holds as ">" for all $\theta_{i+1}\in\varphi'_{i+1}$. In particular, we have

$$V(\theta_{i+1}|\varphi_{i+1},k_{i+1})>V(\theta_{i+1}|\varphi'_{i+1},k'_{i+1}) \text{ for } \theta_{i+1}\in\varphi'_{i+1} \tag{87}$$

Because $E[\theta|\theta\in\varphi'_{i+1}]\geq E[\theta|\theta\in\varphi_{i+1}]$, we have $V(\theta_{i+1}|\varphi'_{i+1},k_{i+1})\geq V(\theta_{i+1}|\varphi_{i+1},k_{i+1})>V(\theta_{i+1}|\varphi'_{i+1},k'_{i+1})$. So we are done by noticing that k_{i+1}improves all types on $\theta_{i+1}\in\varphi'_{i+1}$.
If both (86) and (86) hold, we have

$$V(\theta|\varphi_{i+1},k_{i+1})>V(\theta|\varphi'_{i+1},k'_{i+1}) \tag{88}$$

for all $\theta\in\varphi_{i+1}$ by applying the Monotonicity Lemma to $k_{i+1}<k'_{i+1}$ and (86). As for $\varphi_{i+1'}$, applying the Monotonicity Lemma to (86) and notice that $a_{i+1'}>a'_{i+1}$ because $E[\theta|\theta\in\varphi'_{i+1}]\leq E[\theta|\theta\in\varphi_{i+1'}]$, we have

$$V(\theta|\varphi_{i+1'})>V(\theta|\varphi'_{i+1}) \tag{89}$$

for $\theta\in\varphi_{i+1'}$. (88) and (89) show that the modified contract is better for all types in φ'_{i+1}. As a result, we have a solution that is better than the solution to the modified problem.
To prove the second part, if α'_{i+1} is not $\frac{1}{2}$, we can let $\alpha_{i+1}=\frac{1}{2}$ and at the same time letting k'_{i+1} increase to k_{i+1} so that $V(\overline{\theta}_i)=V(\underline{\theta}_{i+1},k_{i+1}\ ,\alpha'_{i+1}=\frac{1}{2})$. The Monotonicity Lemma implies that $V(\theta,\varphi'_{i+1},k_{i+1},\alpha_{i+1}=\frac{1}{2})>V(\theta,\varphi'_{i+1},k'_{i+1},\alpha'_{i+1})$.for $\theta\in\varphi'_{i+1}$. So we have a better modified solution which is a contradiction. Similar argument shows if $w'_{i+1}(0)>0$, we can increase k'_{i+1} to k_{i+1} so that $V(\overline{\theta}_i)=V(\underline{\theta}_{i+1},k_{i+1},w_{i+1}(0)=0)$. The Monotonicity Lemma yields $V(\theta,\varphi'_{i+1},k_{i+1},w_{i+1}(0)=0)>V(\theta,\varphi'_{i+1},k'_{i+1},w'_{i+1}(0))$.for $\theta\in\varphi'_{i+1}$, which is a contradiction.
Part (1) of Proposition 6 follows directly from Lemma 7. Part (2) follows from Lemma 7 and the Monotonicty Lemma.
We use $t(\varphi(\theta))$ to denote a partition is pooling or separation. Thus the mechanism design problem

is

$$
\begin{aligned}
& \max_{k(\cdot),\Psi,\ t(\cdot)} \sum_i \int_{\varphi_i} V(\theta, k(\varphi(\theta)), t(\varphi(\theta)))f(\theta)d\theta \\
s.t. \qquad & \dot{V}(\theta) = \begin{cases} \frac{r^2}{4c}k(\theta)E[\theta|\varphi(\theta)] \text{ if } t(\varphi(\theta)) = p, \\ \frac{r^2}{4c}k(\theta)\theta \text{ if } t(\varphi(\theta)) = s, \end{cases} && (90) \\
& \dot{k}(\theta) = \frac{-\frac{r^2}{4c}k(\theta)\theta}{\frac{r^2}{4c}\theta^2 - k(\theta)} \text{ if } t(\varphi(\theta)) = s, && (91) \\
& V(\overline{\theta}_i) = V(\underline{\theta}_{i+1}), && (92) \\
& V(1) = \frac{r^4}{32c^2}k(1)E[\theta|c(1)]. && (93)
\end{aligned}
$$

$\dot{V}(\theta) = \frac{d}{d\theta}V(\theta, k(\varphi(\theta)), t(\varphi(\theta)))$ is the growth rate of V. There is no constraints on boundary $\overline{\beta}$. There is also a constraint that $V(\theta) \geq 0$ and $k(\theta) \geq 0$.

Expression (90) is derived from ICs by envelope thoerem. (91) is from the FOC of the firm's IC as in the proof Proposition 5. (93) says that the lowest type's value is determined by $k(1)$ and the information revealed by choosing the contract $c(1)$.

Some observations may be helpful. Intuitively, the problem is essentially find Ψ, $t(\varphi(\theta))$ and $k(1)$. Because if Ψ and t are fixed, for any given $k(1)$, we can determine every $k(\theta)$ and thus $V(\theta)$. More specifically, for any given subinterval with given initial $k(\underline{\theta}_i)$, if $t = s$, we can use (91) to determine all $k(\theta)$ and thus all $V(\theta)$ on that interval. If $t = p$, we have all $k(\varphi_i(\theta)) = k(\underline{\theta}_i)$, we can thus calculate all $V(\theta)$ on the interval. Finally, $k(\underline{\theta}_{i+1})$ can be determined by (92). The problem can thus be solved by choosing the best $k(1) \geq 0$.

Finally, we prove the rest of the Proposition (6). Part (3) is proved by contradiction. Namely, if it is separation at top or bottom, we can improve the firm's expected value.

I. Separation at top.

That is, it is separation on $(\theta_s, \overline{\beta}]$. We will modify the mechanism so that it is pooling on $[\theta_0, \overline{\beta}]$ and everything else is the same. Consider the case where θ_0 is very close to $\overline{\beta}$ so that $\theta_0 > \theta_s$. We will show that for $\theta > \theta_0$, the expected payoff of the firm is higher than in the seperating equilibrium. Denote $\theta_p = E[\theta|\theta > \theta_0]$ and investment for types on $[\theta_0, \overline{\beta}]$ k_p.For the modified mechanism to be incentive compatible, we have

$$V^s(\theta_0) = r^2 k_p \frac{\theta_p}{4c}\theta_0 - \frac{1}{2}k_p^2 \tag{94}$$

In the seperating equilibrium, using Taylor's expansion at θ_0, the firm's payoff is

$$V^s(\theta) = V^s(\theta_0) + \frac{r^2}{4c}k^s(\theta_0)\theta_0(\theta - \theta_0) + \frac{1}{2}\left(\frac{r^2}{4c}\left.\frac{dk^s(\theta)\theta}{d\theta}\right|_{\theta=\theta_0}\right)(\theta - \theta_0)^2 + o[\Delta\theta^2] \tag{95}$$

where $\Delta\theta = \overline{\theta} - \theta_0$. In the proof of Proposition 5, we show that $\frac{\partial V^s(\theta)}{\partial\theta} = \frac{r^2}{4c}k^s(\theta)\theta$ and thus $\frac{\partial^2 V^s(\theta)}{\partial\theta^2} = \frac{r^2}{4c}\frac{dk^s(\theta)\theta}{d\theta}$.

Applying Taylor's Theorem at θ_p, we have

$$
\begin{aligned}
V^s(\theta_0) &= \theta_p k^s(\theta_p)\frac{r^2}{4c}\theta_p - \frac{1}{2}k^s(\theta_p)^2 + \theta_p k^s(\theta_p)\frac{r^2}{4c}(\theta_0 - \theta_p) + \frac{1}{2}\left(\frac{r^2}{4c}\left.\frac{dk^s(\theta)\theta}{d\theta}\right|_{\theta=\theta'}\right)(\theta_p - \theta_0)^2 \\
&= \theta_p k^s(\theta_p)\frac{r^2}{4c}\theta_0 - \frac{1}{2}k^s(\theta_p)^2 + \frac{1}{2}\left(\frac{r^2}{4c}\left.\frac{dk^s(\theta)\theta}{d\theta}\right|_{\theta=\theta'}\right)(\theta_p - \theta_0)^2
\end{aligned}
$$

$\theta' \in [\theta_0, \theta_p]$. Substituting the above equation into (94) and simplifying, we get

$$\frac{1}{2}(\frac{r^2}{4c}\left.\frac{dk^s(\theta)\theta}{d\theta}\right|_{\theta=\theta'})(\theta_p-\theta_0)^2 = \frac{\theta_p k_p r^2\theta_0}{4c}-\frac{1}{2}k_p^2-\theta_p k^s(\theta_p)\frac{r^2\theta_0}{4c}+\frac{1}{2}k^s(\theta_p)^2 = [\theta_p\frac{r^2}{4c}\theta_0-k'](k_p-k^s(\theta_p))$$

$k' \in [\ k_p,\ k^s(\theta_p)]$. The second equality follows from applying Taylor's Theorem to $h(x) = \theta_p x\frac{r^2}{4c}\theta_0 - \frac{1}{2}x^2$. Therefore, $k^s(\theta_p) - k_p = -\kappa(\theta_p - \theta_0)^2$,

$$\kappa = \frac{\frac{1}{2}\left(\frac{r^2}{4c}\left.\frac{dk^s(\theta)\theta}{d\theta}\right|_{\theta=\theta'}\right)}{[\theta_p\frac{r^2}{4c}\theta_0 - k']}$$

Now we show that κ is bounded. To show that, we first notice that $k^s(\theta)\theta$ is continously differentiable, therefore $\frac{dk^s(\theta)\theta}{d\theta}$ is bounded. Second, when θ_0 is close to $\overline{\beta}$, $\theta_p\frac{r^2}{4c}\theta_0 \approx k^*_{\overline{\beta}}$ and $k' \approx k^s(\overline{\beta})$, we know from the proof of Proposition 5 that for a fixed sepration, $k^s(\overline{\beta}) > k^*_{\overline{\beta}}$. Thus κ is indeed bounded for θ_0 close to $\overline{\beta}$.

In the pooling on $[\theta_0, \overline{\beta}]$, the firm's payoff is

$$V^p(\theta) = r^2\frac{\theta_p}{4c}\theta k_p-\frac{1}{2}k_p^2 = V^s(\theta_0)+\frac{r^2}{4c}k_p\theta_p(\theta-\theta_0) = V^s(\theta_0)+\frac{r^2}{4c}k^s(\theta_0)\theta_0(\theta-\theta_0)+\frac{r^2}{4c}(k_p\theta_p-k^s(\theta_0)\theta_0)(\theta-\theta_0)$$

The second equality follows because of (94).

Therefore,

$$\begin{aligned} V^p(\theta) &= V^s(\theta_0) + \frac{r^2}{4c}k^s(\theta_0)\theta_0(\theta-\theta_0) + \frac{r^2}{4c}(k_p\theta_p - k^s(\theta_0)\theta_0)(\theta-\theta_0) \\ &= V^s(\theta_0) + \frac{r^2}{4c}k^s(\theta_0)\theta_0(\theta-\theta_0) + \frac{r^2}{4c}\left[k^s(\theta_p)\theta_p - k^s(\theta_0)\theta_0 - \kappa\theta_p(\theta_p-\theta_0)^2\right](\theta-\theta_0) \\ &= V^s(\theta_0) + \frac{r^2}{4c}k^s(\theta_0)\theta_0(\theta-\theta_0) + \left(\frac{r^2}{4c}\left.\frac{dk^s(\theta)\theta}{d\theta}\right|_{\theta=\theta_0}\right)(\theta_p-\theta_0)(\theta-\theta_0) - o[\Delta\theta^2] \quad (96) \end{aligned}$$

The last equality follows from applying Taylor Theorem to $k^s(\theta_p)\theta_p - k^s(\theta_0)\theta_0$.

Therefore, combining (95) and (96), and taking expectation for $\theta > \theta_0$, we have

$$\begin{aligned} E[V^p(\theta)|\theta &> \theta_0] - E[V^s(\theta)|\theta > \theta_0] \\ &= \left(\frac{r^2}{4c}\left.\frac{dk^s(\theta)\theta}{d\theta}\right|_{\theta=\theta_0}\right)\left\{(E[\theta|\theta>\theta_0]-\theta_0)^2 - \frac{1}{2}E[(\theta-\theta_0)^2|\theta>\theta_0]\right\} + o[\Delta\theta^2] \quad (97) \end{aligned}$$

But because

$$\begin{aligned} E[\theta|\theta &> \theta_0] - \theta_0 = \int_{\theta_0}^{\overline{\theta}}(x-\theta_0)\frac{f(x)}{[1-F(\theta_0)]}dx \\ &= \int_{\theta_0}^{\overline{\theta}}(x-\theta_0)\frac{f(\theta_0)}{[1-F(\theta_0)]}dx + \int_{\theta_0}^{\overline{\theta}}(x-\theta_0)^2\frac{f'(\xi)}{[1-F(\theta_0)]}dx,\ \theta_0 \le \xi \le \overline{\theta} \\ &= \frac{f(\theta_0)}{[1-F(\theta_0)]}\frac{1}{2}(\overline{\theta}-\theta_0)^2 + \frac{f'(\xi)}{[1-F(\theta_0)]}\frac{1}{3}(\overline{\theta}-\theta_0)^3 \\ &= \frac{1}{2}(\overline{\theta}-\theta_0)^2\frac{f(\theta_0)}{f(\theta')(\overline{\theta}-\theta_0)} + \frac{f'(\xi)}{f(\theta')(\overline{\theta}-\theta_0)}\frac{1}{3}(\overline{\theta}-\theta_0)^3,\ \theta_0 \le \theta' \le \overline{\theta} \\ &= \frac{1}{2}(\overline{\theta}-\theta_0)\frac{f(\theta_0)}{f(\theta')} + \frac{f'(\xi)}{f(\theta')}\frac{1}{3}(\overline{\theta}-\theta_0)^2 = \frac{1}{2}(\overline{\theta}-\theta_0) + o(\Delta\theta), \end{aligned}$$

thus we have

$$(E[\theta|\theta > \theta_0] - \theta_0)^2 = \frac{1}{4}(\overline{\theta} - \theta_0)^2 + o(\Delta\theta^2). \tag{98}$$

By the similar way, we can show that

$$E[(\theta - \theta_0)^2|\theta > \theta_0] = \frac{1}{3}(\overline{\theta} - \theta_0)^2 + o(\Delta\theta^2) \tag{99}$$

Substituting (98) and (99) into (97), we have $\left\{(E[\theta|\theta > \theta_0] - \theta_0)^2 - \frac{1}{2}E[(\theta - \theta_0)^2|\theta > \theta_0]\right\} = \frac{1}{4}(\overline{\theta} - \theta_0)^2 - \frac{1}{6}(\overline{\theta} - \theta_0)^2 + o[\Delta\theta^2] = \frac{1}{12}(\overline{\theta} - \theta_0)^2 + o[\Delta\theta^2]$.Therefore, $E[V^p(\theta)|\theta > \theta_0] - E[V^s(\theta)|\theta > \theta_0] > 0$ for $\Delta\theta$ small enough.

II. Separation at bottom

Similar to the top case, we compare it with a pooling at the bottom on $N_\epsilon = [1, 1+\epsilon]$. We let $\theta_0 = 1$ and $\theta_1 = 1 + \epsilon$. We will constuct a pooling so that $V^s(\theta_1) = V^p(\theta_1)$ but the expected value is higher for pooling on N_ϵ. So that the modified mechansim is better because the payoff (and the mechanism) above θ_1 remains the same.

We first let $V^s(\theta_0) = V^p(\theta_0)$, then from the analysis of the top, similar to (**??**), we conclude that

$$V^p(\theta) - V^s(\theta) = v_2\int_{\theta_0}^{\theta} \frac{\theta_1 - \theta_0}{2}dx - v_2\int_{\theta_0}^{\theta}(\theta - \theta_0)dx + \int_{\theta_0}^{\theta} K_p l^2 - K_s(\theta - \theta_0)^2]dx = \frac{v_2}{2}(\theta_1 - \theta)(\theta - \theta_0) + K_0 l^3 \tag{100}$$

where K_0 is constant and $l = \epsilon$. In particular, $V^p(\theta_1) - V^s(\theta_1) = O(l^3)$. We then change $V^p(\theta_1)$ by changing the pooling investment level, so that $V^p(\theta_1) - V^s(\theta_1) = 0$. This is possible because we can do that by changing $V^p(\theta_0)$ by $O(l^3)$. (This is possible because $\max_k V^p(\theta_0) - \max_k V^s(\theta_0) = O(l)$). Then using (**??**), $V_n^p(\theta_1) = V^p(\theta_0) + O(l^3) + \int_{\theta_0}^{\theta_1}(v_1 + \frac{1}{2}v_2 l)d\theta + o(l^3) = V^p(\theta_1) + O(l^3)$. Because we only change $V^p(\theta_0)$ by $O(l^3)$, (100) implies that $V_n^p(\theta) - V^s(\theta) = \frac{v_2}{2}(\theta_1 - \theta)(\theta - \theta_0) + O(l^3)$. Integrate on N_ϵ

$$E[V_n^p(\theta) - V^s(\theta)] = \int_{\theta_0}^{\theta_1}[V_n^p(\theta) - V^s(\theta)]f(\theta)d\theta = \frac{v_2}{12}l^3 + O(l^4) > 0$$

for ϵ small enough.

We have shown that there is no separation at the top, but does it mean that there must be a pooling interval at the top? The answer is not necessarily if there are infinite many partitions. The complication comes from the fact if there are infinite many partitions, there may be an infinite many intervals at the top, with length goes to zero so that they converge to a point, $\overline{\beta}$. Some of those intervals are pooling and some are separation. In this case, there is no top interval so it is meaningless to talk about pooling at the top. Having said that, once there is a top interval, we know it must be pooling. So to prove there is only pooling at the top, we need to rule out infinite partitions clustering at top. The same argument can be made for the bottom. We have ruled out the infinite partitions clustering at top or bottom. We ommit the details here but they are available uppon request.

For part (4), the separation case follows from the proof of Proposition 5 in which we show that there is over-investment with probability 1. The pooling case is proved by contradiction. Suppose first

$$\bar{k}^*_{\varphi_i} < \arg\max_k\{E[V_\theta^\theta|\varphi_i^{p*}]\}. \tag{101}$$

We will show that it is not optimal because increasing $\bar{k}^*_{\varphi_i}$ a bit would improve the mechanism. For the ease of notation, we let $k_i = \bar{k}^*_{\varphi_i}$ and $k_i^* = \arg\max_k\{E[V_\theta^\theta|\varphi_i^{p*}]\}$.

There are two cases. First, increasing k_i would lower $V(\underline{\theta}_i)$. Therefore, increasing k_i by a small amount, Δk, to k_i' would increase the expected firm value on φ_i^{p*}. In particular, it will increase $V(\overline{\theta}_i)$ because (101) implies that k_i is under-investing for type $\theta = E[\theta|\varphi_i^{p*}]$ and thus for $\overline{\theta}_i$ too. This implies that the new investment k_i' would satisfies the modified problem as in stated in Lemma 7 and improve the firm's expected value. A contradiction to Lemma 7.

Second, increasing k_i would increase $V(\underline{\theta}_i)$. In this case, given the quadratic form of $V(\underline{\theta}_i)$ in k_i, there exists a $k_i' > k_i$ such that $V(\underline{\theta}_i)$ remains unchanged (i.e., k_i' and k_i are the two roots). Therefore all types on φ_i^{p*} are better off under k_i' by the Monotonicity Lemma. Since k_i' satisfies the modified problem, this is a contradiction to Lemma 7. If $k_i = k_i^*$, then the solution can be improved in two steps: 1) Increasing k_i to k_i' by a small amount, Δk. Since $k_i = k_i^*$ implies that k_i is over-investing for $\underline{\theta}_i$, it will lower $V(\underline{\theta}_i)$ by an amount of order Δk (it will increase $V(\overline{\theta}_i)$ because k_i is under-investing for $\overline{\theta}_i$). 2) Breaking the φ_i^{p*} into two sub-intervals as in the pooling case of the proof of Lemma 7 (starting from paragraph 4 of the proof), we can improve firm value for all types on φ_i^{p*} by an amount of order Δk. Because step 1 has zero effect on $E[V_\theta^\theta|\varphi_i^{p*}]$ if we ignore higher order terms of Δk and step 2 increases $E[V_\theta^\theta|\varphi_i^{p*}]$ by an amount of order Δk, we improve the firm's expected value with k_i' for small enough Δk. This is a contradiction to Lemma 7 because k_i' satisfies the modified problem as stated in Lemma 7.

References

[1] Ambrose, M. L. and C. T. Kulik (1999) "Old Friends, New Faces: Motivation Research in the 1990s" *Journal of Management* 25, pp. 231-92.

[2] Beaudry, P. (1994) "Why An Informed Principal May Leave Rents To An Agent" *International Economic Review* 34, pp. 821-832.

[3] Benabou R. and J. Tirole (2003) "Intrinsic and Extrinsic Motivation" *Review of Economic Studies*, 70, pp 489-520.

[4] Berkovitch E. and R. Israel (2004) "Why the NPV Criterion Does Not Maximize NPV," *Review of Financial Studies* 17, pp. 239-55.

[5] Bernardo, A., H. Cai, and J. Luo, (2001) "Capital Budgeting and Compensation with Asymmetric Information and Moral Hazard," *Journal of Financial Economics* 61, 311-344.

[6] Bernardo, A., H. Cai, and J. Luo, (2004), "Capital Budgeting in Multi-Division Firms: Information, Agency, and Incentives," *Review of Financial Studies* 17, 739-767.

[7] Bernardo, A., J. Luo, and J. Wang, (2006), "A Theory of Socialistic Internal Capital Markets," *Journal of Financial Economics* 80, 485-509.

[8] Bester, H. and R. Strausz, (2001), "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case," *Econometrica*, vol. 69(4), pages 1077-98.

[9] Burkart, M., D. Gromb, and F. Panunzi, (1997) "Large Shareholders, Monitoring, and the Value of the Firm," The Quarterly Journal of Economics, MIT Press, vol. 112(3), pages 693-728, August

[10] Cho, I. and D. Kreps, (1987), "Signaling games and stable equilibria," Quarterly Journal of Economics, 102(2), pp. 179-222.

[11] Dasgupta, P., P. Hammond, and E. Maskin, (1979), "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *Review of Economic Studies*, 46, pp. 185-216.

[12] Green, J. and J-J. Laffont, (1977) "On the revelation of preferences for public goods," *Journal of Public Economics*, Elsevier, vol. 8(1), pages 79-93, August.

[13] Graham, J. and C. Harvey (2001), "The theory and practice of corporate finance: evidence from the field," Journal of Financial Economics, Elsevier, vol. 60(2-3), pages 187-243, May.

[14] Grinstein, Y. and E. Tolkowsky (2004), "The role of the board of directors in the capital budgeting process – Evidence from S&P 500 Firms". Woring paper.

[15] Gompers P. and J. Lerner (1999), "The Venture Capital Cycle" MIT Press.

[16] Harris, M. and A. Raviv (1996), "The Capital Budgeting Process: Incentives and Information," *The Journal of Finance*, 51, pp. 1139-1174.

[17] Harris, M. and A. Raviv (1998), "Capital Budgeting and Delegation," *Journal of Financial Economics*, 50, pp. 1139.

[18] Hart, O. and J. Moore (1995), "Debt and Seniority: An Analysis in the Role of Hard Claims in Constraining Management," *American Economic Review,* 85, pp. 567-85.

[19] Hermalin, B. (1998), "Toward an Economic Theory of Leadership: Leading by Example," *American Economic Review* 88, pp. 1188-1206.

[20] Holmstrom B. (1982), "Moral Hazard in Teams," *The Bell Journal of Economics*, Vol. 13, No. 2 pp. 324-340

[21] Jensen, M., (1986) "The Agency Costs of Free Cash Flow: Corporate Finance and Takeovers," *American Economic Review*, 76.

[22] Komai M., Stegeman M. and B. Hermalin. (2007) "Leadership and Information." *American Economic Review* 97 (3), 944-947.

[23] Laffont, J-J. and J., Tirole (1988). "The Dynamics of Incentive Contracts." *Econometrica*, 56, 1153-75.

[24] Marino, A.M., and J.G., Matsusaka (2005) Decision processes, agency problems, and information: an economic analysis of budget procedures. *Review of Financial Studies*, 18 pp. 301-325.

[25] Meier, I. and V. Tarhan (2007) "Corporate investment decision practices and the hurdle rate premium puzzle", working paper.

[26] Myers, S. (1977) Determinants of corporate borrowing, *Journal of Financial Economics* 5 pp. 147–175.

[27] Myerson, R. (1979) "Incentive compatibility and the bargaining problem", *Econometrica* 47, 61-73.

[28] Poterba, J. and L. Summers (1995) "A CEO Survey of U.S. Companies' Time Horizons and Hurdle Rates," *Sloan Management Review*, MIT, Fall.

[29] Strausz, R. (2009) "Entrepreneurial Financing, Advice, and Agency Costs," *Journal of Economics and Management Strategy* 18-3 pp. 845-870.