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

15 June 2012

Abstract

This paper considers the optimal design of capital budgeting rules when the firm’s capital budgeting process involves the interaction of top executives, with private information, and stakeholders (e.g., employees), whose motivation and efforts affect the likelihood of the firm’s success. In this setting, the decisions of top executives affect stakeholders’ inferences about firms’ prospects and as a result, influence their actions. Specifically, higher levels of investment expenditures indicate that a firm has promising prospects and induce stakeholders to take actions that contribute to the firm’s success. Within this framework 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, the optimal capital budgeting mechanism is consistent with a number of commonly observed investment distortions such as capital rationing, investment rigidities, overinvestment, and inflated discount rates.

*We would like to thank Adolfo de Motta, David Dicks, Chitru Fernando, Neal Galpin and seminar participants at ANU, DePaul University, 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 Oklahoma, University of Pennsylvania, University of Virginia, UT-Austin, and the XVIII Finance Forum at CEU (Elche-Spain) and the Finance Down Under 2011 Conference at the University of Melbourne. Corresponding author: andres.almazan@mcconbs.utexas.edu.
1 Introduction

The capital investment process has been described (i.e., Brealey, et al. 2011 pp. 241) 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.¹

In this paper, we propose a model of top-down capital budgeting that considers the investment decisions of privately informed top executives whose choices can convey information to the firm’s stakeholders (e.g., lower level managers) whose efforts affect the likelihood of the firm’s success. Within this setting, we examine the optimal design of capital budgeting mechanisms. As we show, the procedures that firms use to evaluate investments influence how the information conveyed by their investment choices is interpreted by stakeholders, and that this in turn influences the capital budgeting rules that executives implement.²

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 such a firm, 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 sources of energy or, alternatively, that it has serious concerns about the future of the more traditional sources of energy.³

¹In 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). See Vayanos (2003) for a model of organizational design that abstracts from incentive considerations but considers the process of information aggregation within the different levels of a hierarchy.

²Supporting 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.”

³In 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/.)
by the investment choice, employees in Shell’s biofuel division may be encouraged to work harder which, in turn, makes the biofuel division more successful.\(^4\)

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. Within this setting, the employee’s effort is more productive when the future prospects of the firm are better which, in combination with the optimal 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 an incentive to overinvest (relative to the case of symmetric information) because doing so conveys favorable information.

Most of our intuition can be illustrated in a setting where the firm’s prospects are one of two types, high or low, and in which the optimal capital budgeting mechanism consists of a menu of two investment-compensation pairs. Within this setting two alternative capital budgeting mechanisms can emerge as optimal: (i) a \textit{separating mechanism} in which a high prospect firm invests more (and compensates the lower manager more) than a low prospect firm and (ii) a \textit{pooling mechanism} in which regardless of their types, firms commit to a fixed level of investment and compensation.\(^5\) With the separating mechanism, 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 mechanism, the firm makes an uninformed investment choice, but the overinvestment tendency is mitigated. Hence, the choice between the pooling and the separating mechanisms is determined by a trade-off between the efficiency gains associated with having an investment policy that incorporates information and the efficiency loss associated with overinvestment.

As we show, the intuition developed in the two type case also holds in the more realistic

\(^4\)The importance of new strategic investments for the firm’s stakeholders is also emphasized in the press release describing Google’s acquisition of Motorola in August 11, 2011. Sanjay Jha, Motorola’s CEO, described the acquisition as “compelling new opportunities for our employees, customers, and partners around the world.”

\(^5\)As we show, a semi-separating mechanism in which there is partial revelation of information is strictly suboptimal in the setting with two types.
case in which the firm’s prospects can be described as coming from a continuum of types. While the analysis of the continuum case raises a number of technical complications, we can show that the optimal mechanism also features investment rigidities (i.e., pooling at the bottom and top of the type distribution) and a tendency toward overinvestment. These findings are important since they confirm that investment distortions are not an artifact of the simple two type model but a necessary feature of the optimal capital budgeting process when investment choices convey information to stakeholders.  

Having established the incentive to overinvest and the potential benefits of committing to a fixed level of investment we enrich our analysis by adding a third party that can independently set capital budgeting policies that influence the optimal mechanism that would be designed by 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 richer setting, we investigate a number of issues that relate 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 committing 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 rate schedule that is a function of investment expenditures.) Third, we consider issues related to executive compensation and in particular EVA-like compensation contracts, i.e, contracts where compensation is determined by the firm’s cash flow minus a pre-specified capital cost.

The analysis shows that several commonly observed practices can enhance firm value within a setting in which investment choices convey information to stakeholders. First, the use of debt financing can create a Myers (1977) debt overhang problem that offsets the incentive to overinvest that arises when the level of investment conveys information to stakeholders. Second, imposing high hurdle rates can increase the opportunity cost of capital for the entrepreneur, offsetting his incentive to overinvest. Finally, compensation

---

6See Section 4 for a precise statement on the implications of the optimal mechanism in the continuum type case.
7See Poterba and Summers (1995) who document the use of high-discount rates on American corporations
contracts can directly affect the entrepreneur’s incentives to use the firm’s investment policy to convey information to the firm’s stakeholders.

As we mentioned at the outset, our analysis of a “top-down” capital allocation process is in contrast to the analysis of “bottom-up” processes that is the focus of the existing literature. Specifically, the existing capital budgeting literature examines the issue of how a firm may distort its capital budgeting practices in order to induce managers to exert proper effort (Bernardo, et al. 2001, 2004, 2006), to curb managers’ empire building tendencies (Harris and Raviv 1996, 1997, Marino and Matsusaka 2005, 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 (Aghion and Tirole, 1997 and Burkart, et al., 1997).

Our contribution also relates to the agency literature that argues 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 being a theory that is consistent with investment rigidities in large corporations our theory is consistent with the high hurdle rates imposed by venture capitalists on the investment choices of young start-up firms, which are unlikely to be subject to these agency issues.8

Our analysis is also closely related to papers in the leadership literature which describe how choices made by leaders influence individuals at lower levels in the organization. Our model is especially close to Hermalin (1998) and Komai, Stegeman, and Hermalin (2007) where an informed leader signals his favorable information by expending greater effort, which in turn motivates his subordinates to work harder (i.e., leadership effects).9 There

---

8Indeed, 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.

9See also Rotemberg and Saloner (2000) for an analysis on how organizations can facilitate innovation by employing a “visionary” CEO who is biased in favor of certain projects. These CEO choices affect employees’ efforts on innovation since employees’ compensation only occur when their ideas are incorporated into the projects chosen by the CEO. By contrast, we consider a profit-maximizing principal who distorts the scale
are, however, a number of key differences between our setting and the setting considered in the leadership papers.

First, these studies focus on how to mitigate undereffort in teams while our analysis concentrates on how the optimal design of capital investment rules can sometimes limit and other times take advantage of leadership effects. Second, 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, by considering contractible investment choices we enrich the contracting space and find that “pooling” tends to be an integral part of the optimal mechanism. Specifically, firms may find it optimal to commit to rigid investment policies, namely investment rules that do not fully incorporate the top executives information. This is in contrast to the leadership papers that focus only on optimal separating mechanisms, i.e., situations in which the leader’s actions fully convey information to the followers.

Moreover, 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 a “budget breaker” further broaden the contract space and provides the entrepreneur with incentives that lead to more efficient investment choices.10

More generally, our paper belongs to the principal-agent literature with an informed principal that faces a moral-hazard problem on the part of the agent. Among other things, this literature considers the effects of the principal’s information on the optimal compensation contract (e.g., Beaudry 1994 and Inderst 2000), the value of the private information to the principal (e.g., Chade and Silvers 2004) and the incentives to disclose information (i.e., provide “advice”) to the agent (e.g., Strausz 2009). In contrast to these signaling models in which the principal makes choices after becoming informed, we consider the possibility that the principal can commit to limit himself to a certain set of alternatives before becoming informed and then chooses among those alternatives after becoming informed.11 Specifically, we consider a mechanism design setting in which the principal commits to a

---

10Holmstrom (1982) first points 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 more efficiently.

11See Biais and Mariotti (2005) for a mechanism design analysis with a similar timing of events.
self-screening mechanism prior to acquiring information and then, after becoming informed, makes choices that convey information that influences the agent’s actions.\textsuperscript{12} Considering a screening rather than a signaling setting not only helps us avoid complications related to multiplicity of equilibria but also provides a natural framework to investigate the design of optimal capital budgeting rules within firms. Furthermore, in contrast to the informed principal without agent’s moral hazard setting analyzed in Maskin and Tirole (1989, 1992), in our setting the revelation principle does not apply. This implies that the design of the optimal (self-screening) mechanisms by the principal cannot be restricted to the set of direct mechanisms and in particular, that mechanisms in which information may be concealed (e.g., pooling mechanisms) can indeed emerge as optimal.

The rest of the paper is organized as follows. Section 2 presents the base model and Section 3 analyzes it. Section 4 considers the continuum of types in the base model. Section 5 modifies the base model by considering a third party and the implementation of corporate budgeting practices and Section 6 presents our conclusions.

2 The model
2.1 The basic set-up

We consider a firm that operates in a risk-neutral economy. The firm is owned and 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.\textsuperscript{13}

The technology of the firm requires investing $I \geq 0$ in order to produce a random amount of output $\tilde{q} \equiv \tilde{z}(e)\sqrt{2I}$. For notational convenience we define $k \equiv \sqrt{2I}$ as the firm’s scale that generates output $\tilde{q} \equiv \tilde{z}(e)k$ at a cost $g(k) \equiv \frac{1}{2}k^2$. Hence, assuming a unitary output price and taking the investment good as the numeraire, the firm’s profit function before the agent’s compensation is:

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

\textsuperscript{12}As Myerson (1983) demonstrates, the revelation principle fails to apply in settings in which the agent takes unobservable actions after the principal reveals his private information.

\textsuperscript{13}In Section 5 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 and hence the design of the optimal capital budgeting mechanism.
As (1) shows, the firm’s profits depend on its scale $k$ and on the agent’s (unobservable) effort $e \in [0, \frac{1}{\beta}]$, which affects the (random) productivity of the invested capital, $\tilde{z}(e)$. Specifically,

$$\tilde{z}(e) = \begin{cases} r & \text{with prob. } \theta e \\ 0 & \text{with prob. } (1 - \theta e) \end{cases}$$  \hspace{1cm} (2)$$

where $r > 0$ and $\theta$ is an exogenous productivity shock described by

$$\theta = \begin{cases} \beta & \text{with prob. } \pi \\ 1 & \text{with prob. } (1 - \pi) \end{cases}$$  \hspace{1cm} (3)$$

where $\beta > 1$. Finally, we assume that the cost of effort to the agent is quadratic in the level of effort $e$ and increases linearly with the firm’s scale $k$: \footnote{We have also investigated a setting in which effort costs are independent of the firm scale $h(e) = \frac{1}{2} c e^2$ but in which investment costs are cubic on firm scale, i.e., $g(k) \equiv k^3$. This alternative setting, while less tractable, produces similar results and is available upon request.}

$$h(e, k) = \frac{1}{2} c e^2 k.$$  \hspace{1cm} (4)$$

The timing of events is as follows: At $t = 0$, before observing $\theta$, the principal offers the agent $w(k, q)$, a compensation schedule contingent on the firm’s investment and the realized output $q$. \footnote{Alternatively, the agent’s compensation schedule can be considered a function of the scale and realized productivity $w(k, z)$. We choose to express it as a function of realized output for expository clarity.} At $t = 1$, the principal privately observes $\theta$ and chooses the investment expenditure $k$. At $t = 2$, the agent makes an unobservable effort choice $e$. At $t = 3$, firm output $\tilde{q}$ is realized and contracts are settled. Figure 1 summarizes the timing of events.

We analyze the model as a mechanism design problem in which the optimal mechanism is determined before the principal becomes informed. In particular, prior to observing $\theta$, the principal commits to choose within a limited set of investment levels and compensation contracts for the agent. \footnote{This specific timing in which the principal (prior to acquiring information) commits to choose from a menu of contracts in also considered by Blais and Mariotti (2005) in their analysis of optimal trading mechanisms in the presence of adverse selection.} Formally, this can be described as a choice at $t = 0$ among a pre-specified set of compensation schedules $w(k, q)$ that are contingent on both the investment...
k chosen at $t = 1$ and the output $q$ realized at $t = 3$.\footnote{A more general schedule would also consider also potential announcements $\hat{\theta}$ by the principal, i.e., $w(k, q, \theta)$. As will later be clear since $k$ can be adapted to the principal’s information, focusing on compensation schedules $w(k, q)$ is without loss of generality.} It is worth mentioning that the mechanism design problem takes into account the effect that the principal’s actions have on the unobservable agent’s effort choices but excludes effort per se as a component of a feasible mechanism. This exclusion is justified since the agent can make effort choices at $t = 2$ that depend on the principal’s choices of $w(k, q)$ and $k$ but the agent cannot commit to a specific level of unobservable effort.

2.2 The symmetric information benchmark

To gain intuition we analyze the case in which both the principal and the agent observe the shock $\theta$ but in which the agent’s effort choice remains unobservable. We denote as $k = \{k_1, k_\beta\}$ and $e = \{e_1, e_\beta\}$ the investment and effort levels that correspond to the two possible realizations of $\theta = \{1, \beta\}$. Without loss of generality, we restrict the analysis to mechanisms in which compensation contracts offer non-negative payments $w = \{w_1, w_\beta\}$ when the realized output is high (i.e., $q \in \{rk_1, rk_\beta\}$) and a zero-payments when the output is low (i.e., $q = 0$).\footnote{In the appendix we show that focusing on compensation contracts with zero rewards when the output is low is without loss of generality.} Thus, the search for the optimal mechanism consists of finding the levels of investment and compensation $m = \{(k_1, w_1), (k_\beta, w_\beta)\}$ and the associated agent’s effort levels $e = \{e_1, e_\beta\}$ that maximize the principal’s value. By denoting $\pi$ and $(1 - \pi)$ as $p_\beta$ and $p_1$ respectively, the principal’s problem can be expressed as

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

s.t.:

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

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

For each realization of the shock, $\theta = \{1, \beta\}$, the principal maximizes firm value (5), subject to the corresponding optimal effort choice (6), and the individual rationality constraint (7). The problem can be simplified because the agent’s limited liability requires $w_\theta \geq 0$ and,
since \( e = 0 \) is feasible, the constraints included in (7) always hold.\(^{19}\)

By substituting the first-order condition of (6) in (5) we get

\[
\max_m V = \sum_{\theta \in \{1, \beta\}} p_\theta \left[ (rk_\theta - w_\theta)\theta \frac{w_\theta \theta}{k_\theta c} - \frac{1}{2} k_\theta^2 \right],
\]

the solution of which leads to the following proposition:

**Proposition 1** In the optimal mechanism \( m^* \) compensation and investment are \( w^*_\theta = \frac{r^2 \theta^2}{8c} \)

and \( k^*_\theta = \frac{r^2 \theta^2}{4c} \) for \( \theta = \{1, \beta\} \). With these choices, the agent’s effort is \( e^*_\theta = \frac{\theta}{2} \) for \( \theta = \{1, \beta\} \) and the principal’s expected payoff is

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

Proposition 1 indicates that in the optimal mechanism \( m^* \), compensation and investment are increasing in \( \theta \), i.e., \( w^*_\beta > w^*_1 \) and \( k^*_\beta > k^*_1 \) and so is the associated agent’s effort i.e., \( e^*_\beta > e^*_1 \). Intuitively, the optimal mechanism results in higher investment and greater induced effort in the more favorable the state because the marginal product of both capital and effort is increasing in \( \theta \). It should also be noted that in the optimal mechanism (i) compensation is a sharing rule independent of \( \theta \), (since \( \frac{w^*_\theta}{r^2 k_\theta} = \frac{1}{2} \)) and (ii) the agent’s effort is increasing in compensation and decreasing in investment \( (e^*_\theta = \frac{w^*_\theta}{r^2 k_\theta}) \).

## 3 Optimal capital budgeting under asymmetric information

The mechanism described in Proposition 1 is not in general optimal when the principal but not the agent observes the shock \( \theta \). In this case, the principal may have an incentive to make an investment choice that is not solely based on its expected productivity but that depends also on the information conveyed about \( \theta \) and its potential influence on the agent’s effort choice.

Technically, when \( \theta \) is privately observed by the principal, the mechanism design problem features a principal that becomes informed (and takes an observable costly action) and an agent who makes an unobservable and privately costly effort choice. In such a framework, the optimal mechanism need not be a direct mechanism in which, after acquiring the private

\(^{19}\)We are implicitly assuming that the entrepreneur is the residual claimant of the output (net of wages) and, thus ignoring mechanisms that consider the possibility of making payments to a third party unrelated to the production process. In Sections 3.1 and 5 we consider how relaxing this assumption affects our results.
information on \( \theta \), the principal has an incentive to truthfully reveal it afterwards, i.e., the revelation principle does not apply. In this setting, direct mechanisms can be suboptimal because while the principal can commit to a specific allocation as a function of the reported \( \theta \), the agent cannot commit to any specific level of unobservable effort after \( \theta \) is reported.\(^{20}\)

In the appendix we show that we can restrict the search for the optimal mechanism to the set of mechanisms that consider two investment-compensation pairs, i.e., \( \{(k_a, w_a), (k_b, w_b)\} \). Within this set, three possible classes of mechanisms can emerge as optimal. If the principal’s choices depend on the observed type, (i.e., the principal chooses \( (k_a, w_a) \) after observing \( \theta = 1 \) and \( (k_b, w_b) \) with \( k_b \neq k_a \) or \( w_b \neq w_a \) after observing \( \theta = \beta \)), then the principal’s choices communicate his private information to the agent. We refer to this class as separating mechanisms. If, however, \( k_a = k_b \) and \( w_a = w_b \), then the principal’s choices convey no information to the agent. We refer to mechanisms in this class as pooling mechanisms. A third possibility consists of mechanisms in which the principal mixes between the two investment-compensation pairs, i.e., partial pooling mechanisms.\(^{21}\)

In what follows, we search for the optimal mechanism within the separating and pooling classes and compare firm value when the optimal mechanisms within these classes 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 separating or the optimal pooling mechanism is considered.

### 3.1 The optimal separating mechanism

In a separating mechanism \( m^s = \{(k^s_1, w^s_1), (k^s_\beta, w^s_\beta)\} \) the principal’s choices of compensation \( w^s = \{w^s_1, w^s_\beta\} \) and investment expenditures \( k^s = \{k^s_1, k^s_\beta\} \) depend on the firm’s type \( \theta \).\(^{22}\)

Furthermore, since \( \theta \) is unobservable to the agent, these choices are subject to incentive compatibility (i.e., “truth-telling”) constraints on the principal’s choices. We define \( V_{\theta}^s \) as

---

\(^{20}\)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, similar to ours, the revelation principle does not apply due to the unobservability of the agent’s effort.

\(^{21}\)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 \( \{\sigma_a, \sigma_b\} \).

\(^{22}\)A positive compensation when realized 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 mechanism, as in the benchmark case, restricting to zero compensation when the output is zero is without loss of generality.
the firm value when the principal observes $\theta$ and chooses $w^*_\theta$ and $k^*_\theta$ for $\theta, \hat{\theta} = \{1, \beta\}$:

$$V^{\hat{\theta}} = (rk^*_{\theta} - w^*_{\theta})\theta e^*_{\theta} - \frac{1}{2}k^*_{\theta}^2$$

(10)

where $e^*_\theta = \arg \max_e\{w^*_{\theta}\theta e - \frac{1}{2}\sigma^2 k^*_{\theta}\}$ for $\hat{\theta} = \{1, \beta\}$ and denote as $e^* = \{e^*_1, e^*_\beta\}$ the corresponding agent’s effort’s vector. Thus, the principal’s problem can be expressed as:

$$\max_{m^*,e^*} V^{\theta} = \sum_{\theta = \{1, \beta\}} p_\theta \left[(rk^*_{\theta} - w^*_{\theta})\theta e^*_{\theta} - \frac{1}{2}k^*_{\theta}^2\right]$$

s.t.:  

$$e^*_\theta = \arg \max_e\{w^*_{\theta}\theta e - \frac{1}{2}\sigma^2 k^*_{\theta}\}, \quad \text{for } \theta = \{1, \beta\}$$

(12)  

$$w^*_{\theta}\theta e^*_{\theta} - \frac{1}{2}\sigma^2 k^*_{\theta} \geq 0, \quad \text{for } \theta = \{1, \beta\}$$

(13)  

$$V^\theta \geq V^{\hat{\theta}}, \quad \text{for } \theta, \hat{\theta} = \{1, \beta\} \text{ and } \hat{\theta} \neq \theta.$$  

(14)

Formally, problem (11)-(14) consists of the addition of the IC constraints (14) to the benchmark problem (5)-(7) where $\theta$ is observable. To solve it, we proceed by first ignoring the IC constraint of the high productivity firm, i.e., $V^\beta \geq V^1_\beta$, and then assuming that the IC of the low productivity firm binds, i.e., $V^1_1 = V^1_\beta$. Then, we substitute it into the objective function (11) and derive the corresponding first order conditions to obtain the optimal separating mechanism, $m^{ss*}$.

**Proposition 2** In the optimal separating mechanism $m^{ss*}$ compensation and investment are: $w^{ss*} = \{w^*_1, \Delta w^*_\beta\}$ and $k^{ss*} = \{k^*_1, \Delta k^*_\beta\}$ where $\Delta = \max \left\{ \frac{1+\sqrt{1-1/\beta^2}}{\beta}, 1 \right\}$. With these choices, the agent’s effort is $e^{ss*}_\theta = e^*_\theta = \frac{r_\theta}{2\alpha}$ for $\theta = \{1, \beta\}$ and the principal’s payoff is

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

(15)

A comparison between Propositions 1 and 2 shows that, relative to the case in which $\theta$ is observable, compensation and investment remain unaltered for the low prospect firm, $(k^*_1, w^*_1) = (k^*_1, w^*_1)$, but increase by the factor $\Delta$ for the high prospect firm, $(k^*_\beta, w^*_\beta) = (\Delta k^*_\beta, \Delta w^*_\beta)$. For both types of firms, however, effort equals the effort exerted in the observable $\theta$ case ($e^*_\theta = e^*_\theta$) since effort depends on the ratio between compensation and investment ($e^*_\theta = \frac{w^*_\theta}{ck^*_\theta}$) which remains unaltered relative to the effort level in the observable case.
The overinvestment factor $\Delta$ summarizes the main intuition of the analysis in the optimal separating mechanism. Specifically, the high prospect firm overinvests (i.e., increases its scale relative to the case of observable $\theta$) to credibly convey to the agent the presence of high prospects. Therefore, the design of the optimal separating mechanism consists of the determination of $\Delta$, namely the minimum required level of overinvestment that makes such communication credible to the agent.

As Proposition 2 indicates, the overinvestment factor $\Delta$ depends on $\beta$, which measures the difference in the size of productivity shocks across firms. Specifically, as a function of $\beta$, the overinvestment factor $\Delta$ has an inverted U-shape that reaches its maximum at $\hat{\beta} = \frac{2\sqrt{3}}{3} \approx 1.15$ and its minimum (i.e., $\Delta = 1$) when $\beta \geq \beta^* \approx 1.84$. In other words, when differences in productivity are relatively small (i.e., when $\beta < \beta^*$) the optimal mechanism requires the high prospect firm to overinvest 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 separating mechanism. When this is the case, the optimal mechanism corresponds to the optimal mechanism in the benchmark case where the agent can observe $\theta$, i.e., $m^{s*} = m^*$. 

Figure 2: Optimal overinvestment factor $\Delta(\beta)$

It is worth stressing that the optimal separating mechanism features overinvestment (and overcompensation) even though it is possible to design a feasible separating mechanism solely based on distortions in compensation. This occurs because, in this setting, investment
distortions can more efficiently convey information about a firm’s prospects. (Since a type 1 firm finds it costlier to overinvest than a type $\beta$ firm whose marginal benefits of a high investment level are higher.) In contrast, a separating mechanism that relies exclusively on distorting compensation is suboptimal because these compensation distortions will be in conflict with the provision of effort incentives to the agent. This is because effort incentives require compensation to be contingent on success, and the probability of success is higher for a type $\beta$ firm. To elaborate, note that a higher $w_\beta$ may help deter the low type from mimicking. But setting a high $w_\beta$ is more costly for the high type than for the low type because the high type is more likely than the low type to pay the high $w_\beta$.

It is also natural to ask whether the principal can employ a mechanism that shapes his own compensation in a way that conveys information about the firm’s type. For example, the principal might consider mechanisms that induce optimal investment and in which information is conveyed by choosing a higher powered principal’s compensation when $\theta = \beta$ than when $\theta = 1$. Note, however, that since the program (11)-(14) includes an adding-up constraint (which implies that the principal is the residual claimant to the firm’s output net of wages) mechanisms that impose distortions in the principal’s payoffs are equivalent to mechanisms that consider distortions in the agent’s compensation. Therefore, our analysis summarized in Proposition 2 does in fact implicitly consider mechanisms in which the principal can shape his own compensation to convey firm’s type information.

We could, however, consider mechanisms in which the adding up constraint is relaxed; i.e., mechanisms that allow the principal to commit to give up a share of the firm’s output in some situations (i.e., to burn money). In the appendix, we show that under a mild parametric condition any mechanism that includes money burning is strictly dominated by the optimal mechanism described in Proposition 2. Intuitively, this occurs because any incentive effect produced by burning money can be replicated by a combination of overinvestment and compensation which, when properly designed, adds some value to the principal. This observation implies, that the optimal separating mechanism derived in Proposition 2 is also the optimal mechanism in a richer setting in which the principal could make positive third-party side-payments or engage in any other form of money burning.24

---

23 Technically, this would relax the adding-up constraint and would make the principal the residual claimant net of wages and any money burned.
24 In the appendix we provide a formal argument of this claim and the specific parametric condition. In Section 5, we revisit this issue by extending the model to include a third party that plays an active role in
3.2 The optimal pooling mechanism

In addition to separating mechanisms, the principal can consider the class of pooling mechanisms in which he commits (before observing $\theta$) to make investment and compensation choices that are independent of the observed $\theta$. Under pooling, the principal avoids the revelation of $\theta$ and thus the agent’s effort choice is independent of $\theta$. We refer to a pooling mechanism as $\tilde{m} = \{\tilde{w}, \tilde{k}\}$ where $\tilde{k}$ and $\tilde{w}$ are, respectively, the investment and the compensation in case of success, and to $\tilde{\theta} \equiv \pi \beta + (1 - \pi)1$ as the average productivity shock.

To find the optimal pooling mechanism $\tilde{m}^*$, the principal solves:

$$\max_{\tilde{m}, \tilde{e}} V_p = (r \tilde{k} - \tilde{w})\tilde{\theta}\tilde{e} - \frac{1}{2}\tilde{k}^2$$  \hspace{1cm} (16)

$$\text{s.t.:} \quad \tilde{e} = \arg\max_{e} \{\tilde{w}\tilde{\theta}e - \frac{1}{2}e\tilde{e}^2\tilde{k}\}. \hspace{1cm} (17)$$

Since the principal’s choices do not convey any information about $\theta$, the agent makes his effort decision based on the average firm type, which leads to the following result:

**Proposition 3** In the optimal pooling mechanism $\tilde{m}^*$ compensation and investment are: $\tilde{w}^* = \frac{\pi^3 \tilde{\theta}^2}{2\epsilon}$ and $\tilde{k}^* = \frac{\pi^2 \tilde{\theta}^2}{4\epsilon}$. With these choices the agent’s effort is $\tilde{e}^* = \frac{\pi \tilde{\theta}}{2\epsilon}$ and the principal’s payoff is $V_p^* = \frac{1}{2}\tilde{k}^*\tilde{e}^*$.

In the optimal pooling mechanism $\tilde{m}^*$, compensation is the same proportion of realized output i.e., $\frac{\tilde{w}^*}{\tilde{k}^*} = \frac{1}{2}$ as in the benchmark and separating optimal mechanisms. Furthermore, $\tilde{m}^*$ induces a level of agent’s effort which is the average of the two effort levels under the benchmark and optimal separating cases, i.e., $\tilde{e}^* = \frac{\pi \tilde{\theta}}{2\epsilon}$. In addition, the investment choice $\tilde{k}^*$ does not depend on the realized $\theta$, which implies that the firm overinvests when $\theta = 1$ and underinvests when $\theta = \beta$. Intuitively, the existence of an optimal pooling mechanism 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.\(^{25}\)

\(^{25}\)In pooling mechanisms firms commit to a maximum investment level. This interpretation of the optimal pooling mechanism 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).
3.3 Separating or pooling mechanism?

In contrast to the optimal separating mechanism, with the optimal pooling mechanism 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. Hence, the choice between the optimal pooling and the optimal separating mechanisms depends on the trade-off between these costs and benefits as the next proposition states.

**Proposition 4**  
*The optimal separating mechanism \( m^{**} \) is more likely to be the overall optimal mechanism: (i) the smaller the likelihood of a high productivity firm (i.e., lower \( \pi \)) and (ii) the larger the difference in productivity among firm types (i.e., higher \( \beta \)).*

Figure 3 displays the regions in the space \((\beta, \pi)\) where each type of mechanism is optimal.

![Figure 3: Comparative statics](image)

As Figure 3 shows, a pooling mechanism is locally optimal, that is, when firms’ productivity levels are sufficiently close (i.e., \( \beta \to 1 \)) the pooling mechanism emerges as optimal. To the extent that firms can exhibit a continuum 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. To gain some intuition on the comparative statics results it is useful to consider some limiting cases. To illustrate the negative effect of \( \pi \) (the probability of a type \( \beta \) firm) on the likelihood of the separating mechanism being optimal we consider the case in which \( \pi \) is close to
1 (i.e., $\theta = \beta$ is highly likely). In this case, for a type $\beta$ firm, the distortion is large within an optimal separating mechanism (the expected overinvestment cost is large) and small under an optimal pooling mechanism (the investment is close to the type $\beta$ full-information investment level).\footnote{For type 1 firms, the investment inefficiency is large under the optimal pooling mechanism and zero under the optimal separating mechanism. However, when $\pi \to 1$ these effects are of second order importance relative to the trade-off that arises for a type $\beta$ firm.}

The positive effect of $\beta$ (the difference in productivity between firm types) on the likelihood of having a separating mechanism as optimal can be easily shown when $\beta$ is large enough and close to $\beta^*$. In this case, the efficiency loss in the optimal pooling mechanism is large because the average investment is far from the full information level for both types. By contrast, the inefficiency inherent in the optimal separating mechanism (namely type $\beta$ overinvestment) diminishes when $\beta$ is sufficiently close to $\beta^*$. In fact, as $\beta$ moves toward $\beta^*$ the optimal separating mechanism converges to the first best mechanism with efficient investment for the type 1 firm and an overinvestment level that approaches zero for the type $\beta$ firm.

4 Optimal capital budgeting mechanisms with a continuum of types

In this section we consider the design of the optimal capital budgeting mechanism in a setting where the productivity shock $\theta$ 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 distortions in optimal capital budgeting. 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 $\theta$ 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 separating mechanism namely the optimal mechanism among those in which different types of firms choose different firm policies. Second, we consider the design of the optimal mechanism among those that allow for the possibility of type bunching. In both cases, to streamline the presentation, we describe the main results of the analysis in the text and place most technical derivations in the appendix.
4.1 Optimal separating mechanism

The search for an optimal separating mechanism consists of finding a schedule of non-negative pairs \( \{(w^s_\theta, k^s_\theta)\}_{\theta \in [1, \tilde{\beta}]} \) such that (i) firms of different types choose different managerial compensation (i.e., bonus) and investment (i.e., \( k^s_\theta \)) if \( \theta \neq \theta' \) and (ii) the schedule of pairs \( \{(w^s_\theta, k^s_\theta)\}_{\theta \in [1, \tilde{\beta}]} \) maximizes expected firm value. We refer to \( w^s_\theta \) and \( k^s_\theta \) as the compensation and investment choices of a type \( \theta \) firm, and to \( e^s_\theta \) as the agent’s induced level of effort under the firm’s choices.\(^{27}\)

Since the agent cannot observe \( \theta \), the firm’s choices within \( \{(w^s_\theta, k^s_\theta)\}_{\theta \in [1, \tilde{\beta}]} \) are subject to incentive compatibility constraints (a type \( \theta \) firm must find it optimal to choose \((w^s_\theta, k^s_\theta)\) rather than any other feasible pair \((w^s_{\theta'}, k^s_{\theta'})\)).\(^{28}\) Specifically, we define \( V^\hat{\theta}_\theta \) as the value of a type \( \theta \) firm that chooses the compensation and investment of a type \( \hat{\theta} \) firm: \[ V^\hat{\theta}_\theta \equiv (rk^s_{\hat{\theta}} - w^s_{\hat{\theta}})\theta e^s_{\hat{\theta}} - \frac{1}{2}k^s_{\hat{\theta}} \] where \( e^s_{\hat{\theta}} = \arg\max_e \{w^s_{\hat{\theta}} e - \frac{1}{2}ee^2k^s_{\hat{\theta}}\} \) for \( \hat{\theta} \in [1, \tilde{\beta}] \). We denote as \( V^\theta \equiv V^\theta_{\hat{\theta}} \) the type \( \theta \) firm value when the firm chooses \((w^s_\theta, k^s_\theta)\), i.e., the pair in the schedule designated for its own type. Thus, the principal’s problem can be expressed as:

\[
\max_{\{w^s_\theta, k^s_\theta, e^s_\theta\}_{\theta \in [1, \tilde{\beta}]} \int_1^{\tilde{\beta}} V^\theta_{\hat{\theta}} f(\theta) d\theta \tag{19}
\]

s.t.:
\[
e^s_\theta = \arg\max_e \{w^s_\theta e - \frac{1}{2}ee^2k^s_\theta\} \quad \text{for} \quad \theta \in [1, \tilde{\beta}], \tag{20}
\]
\[
w^s_\theta \theta e^s_\theta - \frac{1}{2}ee^2k^s_\theta \geq 0 \quad \text{for} \quad \theta \in [1, \tilde{\beta}], \tag{21}
\]
\[
V^\theta_{\hat{\theta}} \geq V^\hat{\theta}_{\theta} \quad \text{for} \quad \theta, \hat{\theta} \in [1, \tilde{\beta}] \text{ and } \theta \neq \hat{\theta}. \tag{22}
\]

The following proposition characterizes the optimal separating mechanism.

**Proposition 5** In the optimal separating mechanism \( \{(w^{*s}_\theta, k^{*s}_\theta)\}_{\theta \in [1, \tilde{\beta}]} \) compensation and investment \( k^{*s}_\theta \) are defined by: \( w^{*s}_\theta = \frac{r}{2}k^{*s}_\theta \) and \( 8ck^{*s3}_\theta - 3\theta^2r^2k^{*s2}_\theta + 4ck^{s3}_1 = 0 \). With these choices, the agent’s effort is \( e^{*s}_\theta = \frac{r^2}{2\theta} \).

\(^{27}\)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 so that it is completely determined by \( w^s_\theta \), a non-negative bonus paid after the positive output is realized.

\(^{28}\)Without loss of generality, the search for the optimal separating mechanism restricts the firm’s choices after observing \( \theta \) to those within the schedule \( \{(w^s_\theta, k^s_\theta)\}_{\theta \in [1, \tilde{\beta}]} \). Such a restriction can be achieved if the firm commits (before observing \( \theta \)) to self-impose a large penalty (e.g., a large wealth transfer) in case it chooses a pair outside the schedule \( \{(w^s_\theta, k^s_\theta)\}_{\theta \in [1, \tilde{\beta}]} \).
As the above proposition indicates, the optimal separating mechanism with a continuum of types resembles the optimal separating mechanism with two types along a number of dimensions. 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., $k^s_{\theta} > k^\sigma_{\theta}$ for $\theta > 1$ and $k^s_1 = k^\sigma_1$). Second, optimal compensation is a sharing rule (i.e., the ratio of compensation to output is constant, $\frac{w^s_{\theta}}{r^s_{\theta}} = \frac{1}{2}$). 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 to compensation. While type non-observability leads to overinvestment and to overcompensation for each type, investment and compensation increase by the same factor and thus their ratio remains unaltered (i.e., $e^s_{\theta} = \frac{w^s_{\theta}}{e^s_{\theta}} = \frac{r^s_{\theta}}{2c} = e^\sigma_{\theta}$).

As we show next, in this setting the optimal mechanism is never characterized by complete separation of types. However, in a different setting, where the timing of events is modified, the equilibrium would be characterized by the previous separating mechanism. Specifically in a signalling framework 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 investment and 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 optimal mechanism described in Proposition 5.\footnote{A similar intuition applies to the case in which $k$ is observable but not contractible.}

4.2 The optimal mechanism

While in the two type case the only alternative to a separating mechanism is a (partial or full) pooling mechanism, with a continuum of types there are other types of mechanisms that consider many other ways in which types can (partially or fully) pool. 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 reduce the search for the optimal mechanism to the class of partition mechanisms which we define next.

A partition mechanism $\Psi^m$ is a mechanism defined by a partition $\Psi = \{\varphi_1, \varphi_2, ...\}$ of the type space $[1, \tilde{\beta}]$ and by a restriction on the (compensation and investment) policies.
followed by the firms whose types belong to a given interval \( \varphi_i \in \Psi \) (i.e., \( \{(w_\theta, k_\theta)\}_{\theta \in \varphi_i}\))\(^{30}\).

Specifically, within each \( \varphi_i \), 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) separating, 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 \( \varphi_i^p \) (respectively \( \varphi_i^s \)) as a subinterval in which firms follow a pooling (respectively separating) policy.

Solving for the optimal \( \Psi^m \) amounts to identifying the partition \( \Psi \) and the schedule of non-negative pairs \( \{(w^c_\theta, k^c_\theta)\}_{\theta \in [1, \beta]} \) associated with \( \Psi \) that maximizes expected firm value. We denote by \( (w^e_\theta, k^e_\theta) \) the compensation and investment choices when \( \theta \in \varphi_i^s \) and by \( (\bar{w}_{\varphi_i}, \bar{k}_{\varphi_i}) \) as the choices when \( \theta \in \varphi_i^p \). We refer to \( e^s_\theta \) and \( \bar{e}_{\varphi_i} \) as the associated efforts and define firm value of a type \( \theta \) firm when choosing the pair of a type \( \hat{\theta} \) firm as:

\[
V_{\hat{\theta}}^\theta = \begin{cases} 
(k^s_{\theta'} - w^e_{\theta'})\theta e^s_{\theta'} - \frac{1}{2}k^2e^s_{\theta'} & \text{if } \hat{\theta} \in \varphi_i^s, \\
(\bar{w}_{\varphi_i} - \bar{e}_{\varphi_i})\theta \bar{e}_{\varphi_i} - \frac{1}{2}k^2\bar{e}_{\varphi_i} & \text{if } \hat{\theta} \in \varphi_i^p.
\end{cases}
\]

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

\[
\max_{\Psi, \{w^c_\theta, k^c_\theta, e^c_\theta\}_{\theta \in [1, \beta]}} E[V_{\hat{\theta}}^\theta] \\
\text{s.t.:} \\
\quad e^s_{\theta} = \max_{e} \{w^e_{\theta'}\theta e - \frac{1}{2}ce^2k^s_{\theta'}\} \quad \text{if } \theta \in \varphi_i^s, \\
\quad \bar{e}_{\varphi_i} = \max_{e} \{\bar{w}_{\varphi_i}\theta |\varphi_i|^p\theta e - \frac{1}{2}ce^2\bar{k}_{\varphi_i}\} \quad \text{if } \hat{\theta} \in \varphi_i^p.
\]

\[V_{\theta}^\theta \geq V_{\hat{\theta}}^\theta, \quad \text{for } \theta, \hat{\theta} \in [1, \beta] \text{ and } \theta \neq \hat{\theta}.\] (27)

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

**Proposition 6** In the optimal mechanism, \( \Psi^m^* = \{\Psi^*, \{(w^*_{\theta}, k^*_{\theta})\}_{\theta \in [1, \beta]}\} \):

1. Compensation is proportional to investment, i.e., \( w_{\theta}^* = \frac{r}{\beta}k_{\theta}^* \).

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

3. Pooling holds at the top and bottom of the type distribution, i.e., \( \exists \ 1 < \theta_L \leq \theta_H < \beta \) such that

30Formally the set \( \Psi = \{\varphi_1, \varphi_2, \ldots\} \) is a partition of \( [1, \beta] \) if (i) \( \bigcup_{i \in \mathbb{N}} \varphi_i = [1, \beta] \), (ii) \( \varphi_i \cap \varphi_j = \emptyset \), and (iii) \( \varphi_i = (\underline{\theta}_i, \overline{\theta}_i] \), including, open, close, half open half close. Notice that the number of intervals \( \varphi_i \) can be countably infinite.
\{ \{ (\bar{\omega}_L^\star, \bar{k}_L^\star) \}_{\theta \in [1, \theta_L]} \cap \{ (\bar{\omega}_H^\star, \bar{k}_H^\star) \}_{\theta \in [\theta_H, \beta]} \} \subset \{ (\bar{\omega}_0^\star, \bar{k}_0^\star) \}_{\theta \in [1, \beta]}.

(4) There is overinvestment in every subinterval of \( \Psi^\star \), i.e., \( k_0^\star \theta > k_0^\star \) if \( \theta \in \phi_i^\star \) and \( \bar{k}_0^\star_i \geq \arg \max_k \{ E[V_\theta^\theta | \phi_i^\star \} \} \) if \( \theta \in \phi_i^\star \).

The previous proposition describes four features of \( \Psi^{m^\star} \) 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., \( \bar{\omega}^\star \frac{\theta}{\bar{k}_0^\star} = \frac{1}{2} \)). Technically, this property simplifies the description of the optimal schedule which is fully characterized by the relationship between investment and the firm’s type i.e., \( k_0^\star \phi \). Second, firm’s investment is monotonically increasing in the firm’s type; that is, more productive firms invest at least as much as their less productive counterparts. Third, \( \Psi^{m^\star} \) 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 and confirms the intuition that rigidities are part of any capital budgeting policy.

Top and bottom pooling in \( \Psi^{m^\star} \) 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.\(^{31}\)

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, et al. 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:

\(^{31}\)For example, numerical analysis (available upon request) shows that a two-pooling-interval partition is optimal if \( \theta \) is uniformly distributed on \([1, 2]\).
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 The practice of capital budgeting: leverage and boards

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 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 which we now enrich 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 can be 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 the project is partially financed with debt from a competitive debt market. Specifically we assume that, before starting operations, the firm issues debt that requires a payment of \( d \) at \( t = 3 \). We consider the following timing of events:

As the previous figure indicates, we assume that debt financing is obtained before the entrepreneur obtains information and consider the mechanism design problem right after the debt has been issued. The firm chooses the amount of debt that maximizes total firm value which, in principle, might require debt in excess of its investment needs (i.e.,
$d > \max\{k_1, k_2\}$). In what follows we first consider the effects of debt financing on the optimal separating mechanism and then briefly discuss the effects of debt financing on the optimal pooling mechanism.

### 5.1.1 Debt and separating mechanisms

To analyze the effect of debt financing on the optimal separating mechanism, we first examine how debt affects the optimal compensation and investment choices in the optimal mechanism and then consider the optimal choice of debt. Formally, for a given debt obligation $d$, we denote by $k^d = \{k_1^d, k_2^d\}$, $w^d = \{w_1^d, w_2^d\}$ and $e^d = \{e_1^d, e_2^d\}$ the investment, compensation and effort exerted when $\theta = \{1, \beta\}$, by $m^d = \{(k_1^d, w_1^d), (k_2^d, w_2^d)\}$ the separating mechanism and by $V_{\hat{\theta}, d} \equiv (rk_\hat{\theta}^d - w_\hat{\theta}^d - d)e_\hat{\theta}^d - \frac{1}{2}k_\hat{\theta}^d$ the principal’s payoff when, after observing $\theta$, offers $w_\theta^d$ and invests $k_\theta^d$ for $\hat{\theta}, \theta = \{1, \beta\}$.

As the equityholder of the firm, the principal solves at $t = 1$:

$$\max_{m^d, e^d} V^d = \sum_{\theta = \{1, \beta\}} p_{\theta}[(rk_{\theta}^d - w_{\theta}^d - d)e_{\theta}^d - \frac{1}{2}k_{\theta}^d]$$ (28)

s.t.:

$$e_{\theta}^d = \arg\max_{e} \{w_{\theta}^d e - \frac{1}{2}ce^2k_{\theta}^d\} \quad \text{for } \theta = \{1, \beta\}$$ (29)

$$V_{\theta, d} \geq V_{\hat{\theta}, d} \quad \text{for } \theta, \hat{\theta} = \{1, \beta\} \text{ and } \hat{\theta} \neq \theta.$$ (30)

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

**Lemma 1** In the optimal separating mechanism, leverage reduces the compensation, investment and induced effort relative to the unlevered case, i.e., $w_{\theta}^{d*} < w_{\theta}^{s*}$, $k_{\theta}^{d*} < k_{\theta}^{s*}$ and $e_{\theta}^{d*} < e_{\theta}^{s*}$ for $\theta = \{1, \beta\}$.

As stated in Lemma 1 the use of debt financing reduces investment for both types of firms in the optimal separating mechanism. This arises from a debt overhang effect that leads to a reduction in investment because the cost of the marginal unit invested is borne by the principal but part of the benefits are realized by the debtholders.

Taking into account the effects of debt described in Lemma 1 for the design of the optimal separating mechanism, we can solve for the debt obligation at $t = -1$ that maximizes firm

---

22
value: \[ \max_d V^d = \sum_{\theta = \{1, \beta\}} p_\theta [(r k^d_{d\theta} - w^d_{d\theta}) e^d_{d\theta} - \frac{1}{2} \lambda^d_{d\theta}]. \] (31)

The solution to this problem is described in the following proposition.

**Proposition 7** If with zero leverage the optimal mechanism is a separating mechanism with overinvestment (i.e., \( \beta < \beta^* \)) then a positive amount of leverage increases firm value \( (d^* > 0) \); if with zero leverage the optimal mechanism is a separating mechanism with efficient investment \( (\beta \geq \beta^*) \) then zero leverage maximizes firm value \( (d^* = 0) \).

This proposition states that debt can improve value if \( \theta \) is unobservable and a separating mechanism is implemented. Intuitively this occurs because debt overhang reduces the incentives of both types to invest, which affects the IC constraints (30), reducing the benefits of low prospect firms of mimicking high prospect firms. In other words, debt overhang decreases the low type’s benefits from overinvesting since overinvestment implies a wealth transfer to the debtholders. This, in turn, reduces the amount the high type must overinvest to separate. As stated in Proposition 7, when the optimal separating mechanism requires overinvestment, the gain obtained by reducing overinvestment for the high prospect firm outweighs the cost due to the underinvestment created on the low prospect firm.\(^{33}\) Of course this is not the case when the optimal separating mechanism features no 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 mechanisms

When choosing within the class of pooling mechanisms, the principal commits (before observing \( \theta \)) to make investment and compensation choices that are independent of the observed \( \theta \). With zero leverage, the principal chooses the pooling mechanism \( \tilde{m}^* = \{\tilde{w}^*, \tilde{k}^*\} \) that maximizes firm value. With positive leverage set at \( t = -1 \), the principal chooses at

---

\(^{32}\) 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.

\(^{33}\) 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.
$t = 0$ the pooling mechanism that maximizes equity value rather than firm value. This occurs because debt is risky regardless of the amount of it raised at $t = -1$, and thus it creates debt overhang and a wedge between firm and equity values. As a result, the optimal pooling mechanism with positive debt reduces firm value relative to the optimal pooling mechanism without debt.

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

From the previous analysis we can make several observations. First, the combination of Propositions 7 and 8 implies that when debt is endogenous, a separating mechanism is more likely to be the optimal mechanism. This is because the possibility of debt financing is worthless when the pooling mechanism is chosen but improves firm value in the case of separating mechanism. Second, the analysis identifies a positive effect of debt financing which has not been considered in the literature so far: When overinvestment is necessary to convey information about a firm’s prospects, debt limits the incentives of low prospect firms to mimic high prospect firms and thus increase firm value. In contrast to other positive effects of debt, this occurs in a setting without taxes and in which firm decisions are made in the best interest of the shareholders.\(^{34}\)

### 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. In either

\(^{34}\) This analysis assumes the firm borrows at $t = -1$ before knowing its type. If the financing choice is made after the entrepreneur knows the firm’s type, the choice may convey information. In such a setting, the low type has an added incentive to mimic the high type, since by doing so it issues overpriced debt. This suggests that in a setting where the financing choices are an integral part of the mechanism design problem, overinvestment by a high prospect firm is likely to be a feature of the optimal separating mechanism.
interpretation, the third party is an entity that has the authority over certain aspects of firm policy but needs to rely on the CEO to obtain information about the investment’s prospects.

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

![Figure 5: Timing of Events](image)

As shown in Figure 5, 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) but then delegates the decisions on agent’s compensation and firm investment to the CEO. In other words, after the initial board action the analysis proceeds as before: the CEO solves a mechanism design problem to maximize his own payoffs.

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 must use to evaluate the firm’s investment choices. To induce the CEO to do this, we assume that the board offers the CEO a compensation contract with “EVA-like” features that induces the CEO to evaluate investment with the discount factor $\delta$. Specifically, the board offers the CEO a share, $\alpha$, 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 $\delta$. 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.

We first examine the case where the mechanism chosen by the CEO is separating, and consider the case of pooling briefly below. We refer to $m^\delta = \{k_1^\delta, w_1^\delta, k_3^\delta, w_3^\delta\}$ as the separating mechanism and denote by $k^\delta = \{k_1^\delta, k_3^\delta\}$, $w^\delta = \{w_1^\delta, w_3^\delta\}$ and $e^\delta = \{e_1^\delta, e_3^\delta\}$ the investment, compensation and effort exerted after $\theta = \{1, \beta\}$ respectively. In addition, we refer to $V_{\hat{\theta}, \delta}$ as the CEO’s payoff when, after observing $\theta$, offers $w_{\hat{\theta}}^\delta$ and invests $k_{\hat{\theta}}^\delta$ for $\hat{\theta}, \theta = \{1, \beta\}$:

$$V_{\hat{\theta}, \delta} \equiv (rk_{\hat{\theta}}^\delta - w_{\hat{\theta}}^\delta)e_{\hat{\theta}}^\delta - \frac{1}{2\delta}k_{\hat{\theta}}^{\delta^2}. \quad (32)$$
Formally, at \( t = 0 \) the CEO solves the following maximization problem:

\[
\max_{m^\delta, e^\delta} V_\delta = \sum_{\theta = \{1, \beta\}} p_\theta [(r k_\theta^\delta - w_\theta^\delta) e_\theta^\delta - \frac{1}{2} \delta^2 k_\theta^2]
\]  \hspace{1cm} (33)

s.t.:

\[
e_\theta^\delta = \arg \max \{w_\theta^\delta e - \frac{1}{2} ce^2 k_\theta^\delta\} \quad \text{for } \theta = \{1, \beta\}
\]  \hspace{1cm} (34)

\[
V_{\theta, \delta} \geq V_{\hat{\theta}, \delta} \quad \text{for } \theta, \hat{\theta} = \{1, \beta\} \text{ and } \hat{\theta} \neq \theta.
\]  \hspace{1cm} (35)

Relative to the basic model without the board, the agent’s problem remains unchanged and the CEO faces the same mechanism design problem as the principal in the basic model except that the investment costs are given by \( \frac{1}{2} \delta^2 k_\theta^2 \) rather than \( \frac{1}{2} \delta^2 k_\theta^2 \). (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 similar steps to those in the analysis in Section 3.1 the solution can be described as follows:

**Lemma 2** Relative to the case in which the discount rate is non-distorted, the optimal separating mechanism \( m^{*\delta} \) scales compensation and investment by \( \delta \) (i.e., \( k_\theta^{*\delta} = \delta k_\theta^{*\delta} \), \( w_\theta^{*\delta} = \delta w_\theta^{*\delta} \)) and leaves the induced agent’s effort unaffected (i.e., \( e_\theta^{*\delta} = e_\theta^{*\delta} \)).

At \( t = -1 \) the board takes into account the distortion that the discount factor \( \delta \) produces on the CEO’s behavior at \( t = 0 \) and, consequently, chooses \( \delta \) to maximize firm value:

\[
\max_{\delta} V^{\delta} = \sum_{\theta = \{1, \beta\}} p_\theta [(r k_\theta^{\delta\delta} - w_\theta^{\delta\delta}) e_\theta^{\delta\delta} - \frac{1}{2} \delta^2 k_\theta^{\delta^2}].
\]  \hspace{1cm} (36)

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

**Proposition 9** If the optimal mechanism is separating with overinvestment when \( \delta = 1 \) (i.e., \( \beta < \beta^* \)) then the board distorts the discount rate upwards (\( \delta^* < 1 \)). If the optimal mechanism is separating without overinvestment (\( \beta \geq \beta^* \)) then the board does not distort the discount rate (\( \delta^* = 1 \)).

To understand the intuition of Proposition 9 it is useful to notice that in problem (36) 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 cash flows are discounted at
higher rates, the optimal separating mechanism solved by the CEO implies a reduction of investment for both type of firms. In particular, similar to the case of debt analyzed above, an infinitesimal reduction from $\delta = 1$ simultaneously diminishes $w_\beta^{\delta}$ and $k_\beta^{\delta}$. Since compensation and investment for the low type ($w_\delta^L$ and $k_\delta^L$) are at their optimal levels, such a reduction has a second order effect on firm value. However, the reduction for the high type ($w_\beta^H$ and $k_\beta^H$), which are above the full information 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 non-distorted.)

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 mechanism (i.e., $\bar{m}^* = \{\bar{w}^*, \bar{k}^*\}$). 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. For this reason, a separating mechanism is more likely to be optimal when the hurdle rate is set by the board since a modified hurdle rate may improve firm value with separation, but does not affect firm value when the CEO chooses among pooling mechanisms.

It should be emphasized that the board creates value by altering the CEO’s objective function which, in turn, can enlarge the set of incentive compatible mechanisms that can be considered. Without the board, the CEO can also design incentive compatible mechanisms that lead to a reduction in investment expenditures. For instance the CEO could commit to transfer output to the agent (or to burn money) when investment expenditures are high. While as shown in the two-party case, these mechanisms are value-reducing and hence suboptimal, the presence of a board can change matters. This is because the board can act as the residual claimant and thus the CEO can make investment choices that are incentive compatible without resorting to additional worker compensation or “money burning”.

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 $\theta$ is observable) since it leads to underinvestment when $\theta = 1$. We conclude this section by considering whether a policy of multiple hurdle rates by the board (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 $\beta$ there exists $\delta^*_\beta$ such that the CEO follows the optimal investment policy: $k^{1*}_1 = k^{s*}_1$ and $k^{2*}_\beta = k^{s*}_\beta$ and its value reaches $V^*$.

Proposition 10 implies that a policy that imposes higher hurdle rates for larger investments eliminates overinvestment when $\theta = \beta$ without inducing underinvestment when $\theta = 1$.$^3$5

5.2.2 Optimal CEO compensation

The analysis in the previous section suggests that it is possible to design a compensation contract for the CEO that fully solves the overinvestment problem. In this section we consider the robustness of this result.$^3$6 In practice there are reasons to be skeptical about the ability of CEO compensation to solve overinvestment distortions in corporations.$^3$7

First, if the board can secretly recontract with the CEO, it has an incentive to secretly induce the CEO to overinvest to convey favorable information to the agent (i.e., the optimality of the separating mechanism considered in the entrepreneurial model would be restored). A second problem can arise if the CEO’s choices require effort, calling for the board to provide the CEO with performance based pay. In such a case, an optimally compensated CEO might find it advantageous to overinvest in order to induce a higher effort from the agent to increase the value of his personal compensation.$^3$8

Finally, although our setting considers a single investment project and a single employee, our intuition can be extended to settings with multiple projects and many stakeholders. In the case of multiple projects, the firm is unlikely to be able to come up with ex ante contracts with its executives that induce them to restore the first best level of investment for each of these projects. We conjecture that in this case the optimal capital budgeting mechanism would impose limits on the executives’ ability to invest in any individual project, (perhaps

---

$^3$5 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.

$^3$6 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 system” by taking advantage of the nonlinearity inherent in the multiple hurdle rate policy.

$^3$7 Notice that this setting also allows the trivial solution of paying the CEO a fixed compensation which would induce the CEO to be indifferent to any allocation (including the first best allocation implemented when $\theta$ is observable).

$^3$8 Indeed we have explored a variation of our setting in which the CEO not only obtains private information but also takes an action that affects the success of the firm. Under some conditions, the optimal mechanism features overinvestment even when CEO compensation is a part of the optimally designed mechanism.
committing to a fixed level of overall corporate investment), and that there would be a tendency to overinvest in those projects with the greatest level of information asymmetry and the greatest need for stakeholder investment or effort.

6 Concluding remarks

Recent survey evidence (e.g., Graham and Harvey 2001) indicates that the discounted cash flow (DCF) approach is widely used by U.S. corporations to evaluate capital investments. While the authors tend to interpret this observation as evidence that academic theory has influenced industry practice, there are still large disconnects between the actual implementation of the DCF approach used in practice and the approach suggested by theory. For example, while the NPV rule proposes discounting expected cash flow estimates with the expected return on investments with equivalent risk, in practice there is a tendency to “inflate” cash flow estimates, and discount then with “inflated” hurdle rates. There is also evidence that firms impose ceilings on the amount that they will invest in their businesses, or alternatively, may require especially high hurdle rates before approving very large capital expenditures.

To better understand this disconnect between theory and practice we propose a “top-down” theory of capital budgeting whose central premise is that the communication process from a firm’s top executives to its lower level employees and the firm’s other stakeholders has a direct influence on the capital budgeting process. This communication process is important because the information conveyed by investment choices can influence stakeholders’ choices, such as employee effort, and these choices can in turn affect firm value. To explore the interaction between investment expenditures and stakeholder choices we first explore an entrepreneurial firm whose production process requires a capital expenditure and effort by its employee. As we show, there is a natural tendency in this setting for firms to overinvest, and because of this tendency, the entrepreneur may commit to rigidities in investment policies, like maximum and minimum levels of investment. We then modify the analysis by introducing a third party (such as an outside financier or a board of directors) who either provides debt financing to the firm or has the authority to set specific capital budgeting policies or can offer compensation contracts to the entrepreneur. Within this setting we find that the use of debt financing and the use high hurdle rates, either imposed by the
third party or induced as part of an EVA compensation contract, can help to offset the overinvestment tendencies that naturally arise when higher investment expenditures convey favorable information.

As we mentioned in the introduction, alternative theories, which are based on the idea that executives enjoy private benefits from investments, also describe roles for a board of director or an outside investor who constrains the investment choices of top executives. In contrast to these theories, our model suggests that there should be a role for investment rigidities in settings with less separation between ownership and control, for example, in situations with VC financed start-up companies. Indeed, our implications are likely to be especially applicable to firms with less tangible assets and greater growth opportunities, i.e., firms in which top executives have private information and where stakeholders play an important role in the firm’s success.

To further distinguish our theory from alternatives it should be noted that there are important differences between the characteristics and the behavior of an ideal board in our setting and the ideal board in a setting where overinvestment arises because of private benefits. Overinvestment that arises because of private benefits can more easily be solved by direct monitoring, suggesting that active boards that closely monitor top executives would be preferred.

In contrast, in our setting, an active board can potentially hinder a firm’s ability to offset the tendency to overinvest. This would be the case if a close relationship makes it easier for the board to offer the executive unobserved (or implicit) side payments that would negate the contractual incentives described in our model that induce the executive to invest less. As a result, while active boards with discretion may be optimal in a setting with private benefits, our setting is consistent with less active, rules-based boards that exercise less discretion.

In a more complicated setting in which both private benefits and incentives to convey information are drivers of managerial overinvestment, boards would find it particularly challenging to mitigate overinvestment. In this case, boards may find it useful to commit to direct investment limits rather than resorting to policies that indirectly constrain managerial investment (e.g., compensation). An analysis of the optimal board actions in this more complicated setting would be quite challenging and will be left to future work.
Appendix: Proofs and other technical derivations

This appendix is divided in three parts: (i) Proofs relative to Sections 2, 3 and 5; (ii) Technical derivations relative to the analysis of the optimal mechanism and (iii) Proofs relative to Section 4.

Part 1: Propositions and results from Sections 2, 3 and 5

Proof of Proposition 1

Let $w_0 = \{w_{1,0}, w_{\beta,0}\}$ be the payments when $z = 0$ and $\theta = \{1, \beta\}$ and, as defined in the text, $w = \{w_1, w_\beta\}$ the payments when $z = r$ and $\theta = \{1, \beta\}$. Thus the principal’s problem is:

$$\max_{k, w_0, w_e} V = \sum_{\theta = \{1, \beta\}} p_\theta \left[(r k_\theta - w_\theta) \theta c_\theta - w_{\theta,0}(1 - \theta c_\theta) - \frac{1}{2} k_\theta^2\right]$$

s.t.: 

$$e_\theta = \arg \max_e \{w_\theta \theta e + w_{\theta,0}(1 - \theta e) - \frac{1}{2} c e^2 k_\theta\}, \quad \text{for } \theta = \{1, \beta\}$$

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

First, 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 $w_\theta$ while keeping $(w_\theta - w_{\theta,0})$ constant increases firm value without affecting the agent’s effort incentives. Second, we impose $w_{\theta,0} = 0$, and obtain $w_\theta^* \text{, } e_\theta^* \text{ and } k_\theta^*$ by first substituting the first order condition of $e_\theta$ (38) and then solving in the first order conditions of $w_\theta$ and $k_\theta$. \]

Proof of Proposition 2

If $\beta \geq \beta^*$ then the unconstrained solution described in Proposition 1 satisfies $V_\beta^1 \geq V_\beta^1$ (i.e., $IC_\beta$ and $V_1^1 \geq V_1^1$, 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 IC1 which requires us to solve the general problem case in which $w_{\theta,0}^* > 0$. Let $w_0^* = \{w_{1,0}^*, w_{\beta,0}^*\}$ be the wage when $z = 0$ and $\theta = \{1, \beta\}$. Ignoring (by now) IC$\beta$ we get:

$$\max_{w_0^*, w_e, k^*, \epsilon^*} (1 - \pi)V_1^1 + \pi V_\beta^1$$

s.t.: 

$$e_\theta^* = \arg \max_e \{w_\theta^* \theta e + w_{\theta,0}^*(1 - \theta e) - \frac{1}{2} c e^2 k_\theta^*\}, \quad \text{for } \theta = \{1, \beta\}$$

$$w_\theta^* \theta e_\theta^* - \frac{1}{2} c e^2 k_\theta^* \geq 0, \quad \text{for } \theta = \{1, \beta\}$$

$$V_1^1 \geq V_1^\beta$$

where $V_\beta^\beta = rk_\beta^* e_\beta^* - [w_{\beta,0}^* + (w_\beta^* - w_{\beta,0}^*) \theta e_\beta^*] - \frac{1}{2} k_\beta^2$.

Notice that in the previous problem, type 1’s payoff is maximized as in the unconstrained case (i.e., $w_{1,0}^* = 0$, $w_{\beta}^* = w_{\beta,0}^*$, and $k_{1}^* = k_1^*$) because any deviation (i.e., $w_{1,0}^* \neq 0$, $w_{\beta}^* \neq k_1^*$, or $k_{1}^* \neq k_1^*$) reduces $V_1^1$ without easing IC1. We impose such values and solve for type $\beta$’s optimal values:

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

s.t.: 

$$e_\theta^* \in \arg \max_e \{w_{\theta}^* + (w_\theta^* - w_{\theta,0}^*) \beta e - \frac{1}{2} c e^2 k_{\beta}^*\}, \quad \text{for } \theta = \{1, \beta\}$$

$$V_1^1 \geq V_1^\beta.$$
Expression (46) can be rewritten as:

\[
V_1 = \frac{1}{2}k_1^2 \geq \left\{ rk_\beta^s - (w_\beta^s - w_\beta^s, 0) \right\} e_\beta^s - \frac{1}{2}k_\beta^s - w_\beta^s, 0 = V_1^\beta. \tag{47}
\]

We define \( \alpha_\beta \equiv \frac{w_\beta^s - w_{\beta, 0}}{r_k^s} \), express (45) as \( e_\beta = \frac{\alpha_\beta r_\beta^s}{c} \) and plug them into the objective function:

\[
\max_{w_\beta^s, k_\beta^s, \alpha_\beta} (1 - \pi) \frac{k_1^2}{2} + \pi \left\{ rk_\beta^s (1 - \alpha_\beta) \beta^2 \alpha_\beta r_\beta^s \frac{c}{c} - \frac{1}{2}k_\beta^s - w_\beta^s, 0 \right\},
\]

whose Lagrangian is:

\[
L = \frac{(1-\pi)k_1^2}{2} + \pi \left\{ \frac{rk_\beta^s (1 - \alpha_\beta) \beta^2 \alpha_\beta r_\beta^s}{c} - \frac{k_\beta^s}{2} - w_\beta^s, 0 \right\} - \lambda \left\{ \frac{rk_\beta^s (1 - \alpha_\beta) \alpha_\beta r_\beta^s}{c} - \frac{k_\beta^s}{2} - w_\beta^s, 0 - \frac{k_1^2}{2} \right\} + \mu w_\beta^s, 0.
\tag{48}
\]

By Kuhn-Tucker Theorem, the FOCs with respect to \( w_\beta^s, 0, \alpha_\beta \), and \( k_\beta^s \) are:

\[
\lambda - \pi + \mu = 0 \tag{49}
\]

\[
(\pi \beta - \lambda)(1 - 2\alpha_\beta) = 0 \tag{50}
\]

\[
(\pi \beta - \lambda) \alpha_\beta r_\beta^s \beta c^{-1} - (\pi - \lambda)k_\beta^s = 0 \tag{51}
\]

Since by (49), \( \pi \beta - \lambda = \mu + (\beta - 1)\pi > 0 \), then (50) implies \( \alpha_\beta = \frac{1}{2} \) and (51) \( (\pi - \lambda)k_\beta^s > 0 \). Thus \( \pi - \lambda = \mu > 0 \) which implies \( w_\beta^s, 0 \geq 0 \) is binding. Equation (51) implies that \( k_\beta^s = \frac{(\pi \beta - \lambda)r(1 - \alpha_\beta) \alpha_\beta r_\beta^s}{c} > r(1 - \alpha_\beta) \alpha_\beta r_\beta^s = k_\beta^s \) (i.e., type \( \beta \) overinvests). Solving for \( k_\beta^s \) in the binding (47) after substituting \( \alpha_\beta = \frac{1}{2} \) and \( w_\beta^s, 0 = 0 \), we get

\[
k_\beta^s = \frac{r_k^s r_\beta^s}{4c} - \frac{1}{2}k_\beta^s - \frac{k_1^2}{2} = 0 \tag{52}
\]

which has two real roots. But since \( V_2^\beta = \beta r_k^s r_\beta^s - \frac{1}{2}k_\beta^s - \frac{1}{2}k_\beta^s r_\beta^s (\beta - 1) + r_k^s r_\beta^s - \frac{1}{2}k_\beta^s \), we get:

\[
V_2^\beta = rk_\beta^s r_\beta^s (\beta - 1) + \frac{k_1^2}{2} \tag{53}
\]

which follows from (52). Since \( V_2^\beta \) is increasing in \( k_\beta^s \), the larger root is optimal i.e., \( k_\beta^s = k_\beta^s + \sqrt{r_k^s r_\beta^s - k_\beta^s} \). Substituting \( k_\beta^s = \frac{r_k^s r_\beta^s}{4c} \), we get \( k_\beta^s = \frac{1 + \sqrt{1 - 1/\beta}}{\beta}k_\beta^s \). Notice that \( \frac{1 + \sqrt{1 - 1/\beta}}{\beta} = 1 \), but two real roots, 1, and \( \beta^* \). However, since (i) \( \beta^* \) is the largest real root, (ii) \( \lim_{\beta \to \infty} \frac{1 + \sqrt{1 - 1/\beta}}{\beta} < 1 \) and (iii) \( \frac{1 + \sqrt{1 - 1/\beta}}{\beta} > 1 \) for \( \beta \) close to 1, then \( \frac{1 + \sqrt{1 - 1/\beta}}{\beta} > 1 \) for \( 1 < \beta < \beta^* \). Finally, to check IC, notice that (53) and \( V_1 = k_1^2 \) implies that IC3 holds if \( k_\beta^s \beta > k_1^s \) which follows since \( \beta > 1 \) and \( k_\beta^s > k_\beta^s > k_1^s \). The principal’s payoff (15) follows from plugging in the objective function.

**On the optimality of money burning mechanisms**

We will show that money burning arrangements are not optimal if and only if

\[
c \leq \frac{\Delta \beta^2 r}{2} \tag{54}
\]

32
which is equivalent to \( k_{\beta}^{ss} \leq \frac{c}{\theta} \), i.e., the marginal benefit of the high type firm from investing conditional on success is higher than the marginal cost.

Consider a separating mechanism \( m \) include a third party payment. There are three possibilities of payment. \( g \geq 0 \) before investing so that \( k_{\beta} = k_{\beta}^s - g \), after output success \( r \cdot k_{\beta} \), \( g_r \geq 0 \) and after failure \( 0 \). By building the Lagrangian as in (48) and taking first order condition, we get \( g = 0 \) and \( g_r = 0 \). Intuitively, \( g \) and \( g_r \) payments cost the high type more and which make separation more costly. The FOCs (49)-(51) in addition to the FOC for \( g_r \) gives

\[
(\pi_\beta - \lambda)e_\beta - (\pi - \lambda) + v = 0
\]

where \( e_\beta = \alpha_\beta \beta \) and \( v \geq 0 \) is the multiplier for \( g_r \geq 0 \). (51) is

\[
(\pi_\beta - \lambda)e_\beta (1 - \alpha_\beta) r - (\pi - \lambda) k_{\beta}^s = 0.
\]

Substituting the optimal \( \alpha_\beta = \frac{1}{2} \) (which follows as in the proof of Proposition 2), we get from (55) and (56)

\[
(\pi - \lambda)(\frac{k_{\beta}^s}{r/2} - 1) = -v
\]

There are two cases. First \( v \geq 0 \), (57) implies \( k_{\beta}^{ss} \leq \frac{c}{2} \) because (56) implies \( \pi - \lambda > 0 \) (otherwise \( e_\beta \leq 0 \), which is either impossible or clearly not optimal). In this case, \( g_r = 0 \) and the solution is the same as that characterized in Proposition 2. In particular, \( k_{\beta}^s = k_{\beta}^{ss} \leq \frac{c}{2} \), which is satisfied if (54) holds. On the other hand if (54) does not hold, we will have have \( g_r > 0 \). Suppose not. We have \( g_r = 0 \) and \( k_{\beta}^s = k_{\beta}^{ss} \) and thus \( \frac{k_{\beta}^s}{r/2} - 1 > 0 \) because (54) does not hold. This implies that the left hand side of (57) is positive, which implies \( v < 0 \), a contradiction. In summary, if (54) holds, our focus on the set of mechanisms considered in (11)-(14) is w.l.o.g.

**Proof of Proposition 3**

It follows directly from taking the FOC in the program (16)-(17)

**Proof of Proposition 4**

If \( \beta \geq \beta^* \) a separating mechanism dominates. By contrast, when \( \beta < \beta^* \) a pooling mechanism dominates if and only if \( V_p^* > V_s^* \equiv k_{\beta}^2 \cdot 2 + \pi(1 - \frac{1}{\beta}) \Delta k_{\beta}^2 \), that is when \( \frac{k_{\beta}^2}{2} - \frac{k_{\beta}^2}{2} \geq \pi(1 - \frac{1}{\beta}) \Delta k_{\beta}^2 \).

Substituting \( k_{\beta}^* = \frac{c^2}{\theta^2}, k_{\beta}^s = \frac{c^2}{\theta^2} \) and \( \Delta = \frac{\beta + \sqrt{\beta^2 - 1}}{\beta} \), we get \( \frac{[1+\pi(1-1)]^4}{2} - \frac{1}{2} \geq \pi(1 - \frac{1}{\beta}) \beta^2(\beta + \sqrt{\beta^2 - 1}) \cdot \frac{1}{2} \geq 0 \).

To sign \( \zeta \), fix \( \beta \) and let \( \pi^*(\beta) \) such that \( \zeta = 0 \). Notice that \( \frac{dK}{d\pi} \bigg|_{\pi=0} = (\beta - 1)\beta(\beta + \sqrt{\beta^2 - 1}) \geq 0 \).

If \( \beta > \frac{2}{\sqrt{3}} \), then \( \frac{dK}{d\pi} \bigg|_{\pi=0} < 0 \) and thus \( \zeta < 0 \) for \( \pi \) close to 1. Also \( \zeta(\pi = 1) > 0 \) because \( V_p^*(\pi = 1) = \frac{k_{\beta}^2}{2} \) achieves the full information payoff. Thus by continuity, \( \exists \pi^*(\beta) > 0 \) such that \( \zeta = 0 \). Further, since \( \zeta(\pi) \) is convex, \( \zeta(\pi) \leq 0 \) for any \( 0 \leq \pi \leq \pi^*(\beta) \). In addition, because

\[
0 = \zeta(\pi^*) - \zeta(0) = \int_0^{\pi^*} \frac{dK}{d\pi}(\pi)d\pi < \int_0^{\pi^*} \frac{dK}{d\pi}(\pi^*)d\pi = \frac{dK}{d\pi}(\pi^*)\pi^*, \text{ then } \frac{dK}{d\pi} \bigg|_{\pi=\pi^*(\beta)} > 0.
\]

That is, \( \frac{dK}{d\pi} > 0 \)
for all π > π*(β) by convexity and ζ > 0 for all π > π*(β). Thus (i) holds if π > π*(β).
Consider ζ(β, π*(β)) = 0. Simple algebra shows \( \frac{\partial \pi^*}{\partial \beta} \bigg|_{\pi = \pi^*(\beta)} < 0 \) and the Implicit Function Theorem applied to ζ(β, π*(β)) gives: \( \frac{\partial \pi^*}{\partial \beta} \bigg|_{\pi = \pi^*(\beta)} > 0 \). Thus, π*(β) is strictly increasing on (\( \frac{2}{\sqrt{3}}, \beta^* \)). Let \( \beta^{**}(\pi) \): (0, 1) → (\( \frac{2}{\sqrt{3}}, \beta^* \)) be the inverse function of π*(β). For a fixed π, ζ(β) < 0 if \( \beta^{**}(\pi) < \beta < \beta^* \). Now since ζ(β1, π*(β1)) = 0 and π*(β) > 0, if \( \beta_1 > \beta^{**} \) then \( \pi^*(\beta_1) > \pi^*(\beta^{**}) = \pi \) and thus separating is optimal for (π, β1) by part (i). The case where \( \beta_1 < \beta^{**} \) is similarly proved.

**Proof of Lemma 1**

As in the proof of Proposition 2: (i) \( w^d_0 = \frac{rk^{d^*}_1 - d}{c} \); (ii) wage is zero when the output is zero regardless of whether IC1 is binding or not, (iii) only IC1 can bind and (iv) \( V_1^{1,d} \) reaches the full information payoff. As in the proof of Proposition 1, for type 1: \( (w^{d,1}_1, k^{d,1}_1, e^{d,1}_1) = \left( \frac{r k^{d^*}_1 - d}{c d}, \frac{r^2}{4 c k^{d^*}_1}, \frac{w^{d^*}_1}{c k^{d^*}_1} \right) \).

Since \( k^{d,1}_1 < k^{d^*}_1 \), it follows that Lemma 1 holds for type 1.

Next we show that IC1 doesn’t bind at \( k^{d}_1 = k^{d^*}_1 \) and \( k^d_1 \) that maximizes \( V^{1,d}_1 \). When \( d = 0 \), IC1 is:

\[
\left( \frac{r k^{d^*}_1}{c} \right)^2 \ \

\beta - \frac{1}{2} k^{d^*}_1 < \left( \frac{r k^{d^*}_1}{c} \right)^2 \ \

\beta - \frac{1}{2} k^{d^*}_1 \leq \left( \frac{r k^{d^*}_1}{c} \right)^2 \ \

\beta - \frac{1}{2} k^{d^*}_1 \leq V^{1,d}_1.
\]

But \( \left( \frac{r k^{d^*}_1}{c} \right)^2 \beta - \frac{1}{2} k^{d^*}_1 = \int_0^1 \left( \frac{r k^{d^*}_1}{c} \right)^2 \beta - \frac{1}{2} k^{d^*}_1 dx \)

and \( \left( \frac{r k^{d^*}_1}{c} \right)^2 \beta > \left( \frac{r k^{d^*}_1}{c} \right)^2 \beta \) because \( k^{d^*}_1 > k^{d^*}_1 \). Therefore:

\[
\left( \frac{r k^{d^*}_1}{c} \right)^2 \beta - \frac{1}{2} k^{d^*}_1 < \left( \frac{r k^{d^*}_1}{c} \right)^2 \beta - \frac{1}{2} k^{d^*}_1 \leq V^{1,d}_1.
\]

The second inequality follows because \( k^{d^*}_1 \) may not be optimal when \( d > 0 \). Expression (59) is IC1 under \( d > 0 \) after substituting \( w^d_0 = \frac{rk^{d^*}_1 - d}{c} \). Since \( k^{d^*}_1 \) is (weakly) overinvestment for type β when \( d = 0 \), then there is strictly overinvestment for type β when \( d > 0 \) (i.e., under full information \( k^{d^*}_1 = \left( \frac{2}{\sqrt{3}} - \frac{d^2}{4 c k^{d^*}_1} \right) \beta^2 < k^{d^*}_1 \)). Since \( k^{d^*}_1 \) satisfies IC1 with strict inequality, \( k^{d^*}_1 < k^{d^*}_1 \) is optimal because it reduces overinvestment. Finally \( w^{d^*}_0 = w^{d^*}_0, e^{d^*} < e^{d^*} \) follows from \( w^d_0 = \frac{rk^{d^*}_1 - d}{c} \) and \( e^{d^*}_0 = \frac{w^{d^*}_0}{c k^{d^*}_1} \).

**Proof of Proposition 7**

Taking derivatives: \( \frac{\partial V}{\partial d} \bigg|_{d=0} = \frac{\partial V}{\partial k^d_1} \bigg|_{d=0} \frac{dk^d_1}{d} \bigg|_{d=0} \). Since by Lemma 1 \( \frac{dk^d_1}{d} \bigg|_{d=0} < 0 \) and by assumption

\[
\frac{\partial V}{\partial d} \bigg|_{d=0} < 0 \text{ then } \frac{\partial V}{\partial d} \bigg|_{d=0} > 0. \text{ The second part of the proposition is immediate.}
\]

**Proof of Proposition 9**

First, we show that \( \delta = 1 \) is suboptimal.

Let: \( \frac{\partial V_1}{\partial \delta} = \frac{\partial V_1}{\partial k^d_1} \frac{dk^d_1}{d} + \frac{\partial V_1}{\partial k^d_1} \frac{dk^d_1}{d} + \frac{\partial V_1}{\partial k^d_1} \frac{dk^d_1}{d} + \frac{\partial V_1}{\partial k^d_1} \frac{dk^d_1}{d} \). Since \( \alpha_\delta \) satisfies the FOC \( \frac{\partial V_1}{\partial k^d_1} \bigg|_{\alpha(\theta) = \frac{1}{2}} = \frac{\partial V_1}{\partial k^d_1} \bigg|_{\delta=1} = 0 \). Similarly, \( \frac{\partial V_1}{\partial k^d_1} = 0 \) at \( \delta = 1 \), since \( k^d_1 \) satisfies FOC \( \frac{\partial V_1}{\partial k^d_1} \bigg|_{k^d_1=k^d_1} = \frac{\partial V_1}{\partial k^d_1} \bigg|_{\delta=1} = 0. \)
Finally, \( \frac{\partial V}{\partial \delta} \bigg|_{\delta=1} = 0 \) because \( k^*_\beta > k_\beta^* \). Therefore, since \( \frac{dk^*_\beta}{ds} > 0 \), \( \frac{\partial V}{\partial \delta} \bigg|_{\delta=1} = \frac{\partial V}{\partial s} \bigg|_{s=1} = \frac{\partial V}{\partial \delta} \bigg|_{\delta=1} < 0 \). That is, decreasing \( \delta \) increases \( V_\delta \) at \( \delta = 1 \).

Second, we show that \( \delta > 1 \) is suboptimal. Notice that the optimal contract under \( \delta > 1 \), \( (w^\delta_\theta, k^\delta_\theta) \), satisfies IC\(_1\) under \( \delta = 1 \). When \( \delta > 1 \), IC\(_1\) is

\[
\delta^4 k^*_\beta r(1 - \frac{1}{2})e^*_\beta - \delta^4 \frac{k^*_\beta - 2}{2} \leq \delta^4 k^*_1 r(1 - \frac{1}{2})e^*_1 - \delta^4 \frac{k^*_1 + 2}{2}.
\]

Plugging in the optimal \( (w^\delta_\theta, k^\delta_\theta) \) when \( \delta = 1 \), the IC\(_1\) becomes

\[
\delta^3 k^*_\beta r(1 - \frac{1}{2})e^*_\beta - \delta^4 \frac{k^*_\beta - 2}{2} \leq \delta^3 k^*_1 r(1 - \frac{1}{2})e^*_1 - \delta^4 \frac{k^*_1 + 2}{2}.
\]

Comparing both inequalities, we find that subtracting \( \delta^3 k^*_\beta r(\delta - 1)e^*_\beta \) from the LHS and \( \delta^3 k^*_1 r(\delta - 1)e^*_1 \) from the RHS of (60) yields (61). But

\[
\delta^3 k^*_\beta r(\delta - 1)e^*_\beta > \delta^3 k^*_1 r(\delta - 1)e^*_1,
\]

iff \( \delta > 1 \). Therefore, (61) is satisfied with strict equality. We can also check that \( (w^\delta_\theta, k^\delta_\theta) \) satisfies all the constraints in program (44) to (46). As a result, \( (w^\delta_\theta, k^\delta_\theta) \) is strictly suboptimal under \( \delta = 1 \), since at the optimum IC\(_1\) must bind. Hence, firm value when under \( (w^\delta_\theta, k^\delta_\theta) \) is less than \( V_{k=1} = V^* \). Technically, reducing \( \delta \) from 1 relaxes IC\(_1\) constraint, which is binding at \( \delta = 1 \). Notice that the inequality in (62) flips for \( \delta < 1 \). Therefore, (61) is more difficult to satisfy than (60).

**Part 2: The optimal mechanism (technical derivations)**

The results in Proposition 4 confirm that the Revelation Principle (Laffont and Green, 1977, Myerson, 1979, and Dasgupta, Hammond, and Maskin, 1979) does not hold in this setting. Since the agent’s effort choice maximizes the agent’s payoff conditional on the principal’s announcements, solving for the optimal mechanism requires an additional incentive compatibility constraint which breaks the equivalence between a direct mechanism, in which the principal announces the type, and the indirect mechanism in which the principal does not. As it turns out, the optimal mechanism features either full separation (i.e., a full disclosure of the private information) or full pooling (i.e., a complete absence of disclosure). To prove this property, we first show that w.l.o.g. the mechanism design problem can be solved with a menu of two contracts. Specifically, we denote a contract as \( c_i \equiv (w_i(q), k_i) \) and consider a general menu of contracts \( \{c_i\}_{i \in I} \) where \( I \) may be infinite (and through which types may not be fully revealed). Let \( \psi(\theta) \) be the set of contracts that \( \theta \) chooses with positive probability. Let \( w_{i,0} \equiv w_i(0) \) and \( w_i \equiv w_i(rk_i) \) and let the probability of each type choosing \( c_i \), be \( \pi_\theta(c_i) \) for \( c_i \in \psi(\theta) \). Then the principal’s problem is

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

s.t.: \[
\begin{align*}
&c_i \in \arg \max_e \{w_{i,0} + (w_i - w_{i,0})E[\theta|c_i] - h(e, k_i)\}, \\
&w_{i,0} + (w_i - w_{i,0})E[\theta|c_i] - h(e, k_i) \geq 0, \\
&V_\theta(c_i) \geq V_\theta(c_j), \text{ for any } c_i \in \psi(\theta) \\
&\sum_i \pi_\theta(c_i) = 1, \\
&w_i \geq 0, w_{i,0} \geq 0.
\end{align*}
\]
We define \( \alpha_i \equiv \frac{w_{i,j} - w_{i,0}}{k_{i,r}} \) and \( a_i \equiv r k_i (1 - \alpha_i) e_i \) and state the following monotonicity result:

**Lemma 3 (Monotonicity)** Any feasible mechanism requires that: (i) If \( \theta > \theta' \) picks \( c_i \) with positive probability and \( \theta' \) picks \( c_j \) with positive probability, then \( a_i \geq a_j \); and (ii) If \( \theta > \theta' \) weakly prefers \( c_j \) to \( c_i \) with positive probability, \( \theta' \) strictly prefers \( c_j \) to \( c_i \); if \( \theta' < \theta \) weakly prefers \( c_i \) to \( c_j \) with \( a_i > a_j \), then \( \theta \) strictly prefers \( c_i \).

**Proof.** (i): IC constraints imply: \( a_i \theta - \frac{1}{2} k_i^2 - w_{i,0} \geq a_j \theta - \frac{1}{2} k_j^2 - w_{j,0} \) and \( a_j \theta' - \frac{1}{2} k_j^2 - w_{j,0} \geq a_i \theta' - \frac{1}{2} k_i^2 - w_{i,0} \). Adding them up: \( (a_i - a_j)(\theta - \theta') \geq 0 \).

(ii): \( \theta \)'s preference implies: \( \frac{1}{2} k_i^2 - \frac{1}{2} k_j^2 + w_{i,0} - w_{j,0} \geq \theta (a_i - a_j) \) which implies that \( \theta' \) strictly prefers \( c_j \). The other case follows by a similar argument.

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

**Proof:** We distinguish two cases:
Case 1: Any multi-contract mechanism with \( \theta \in \{1, \beta\} \) in which \( \theta \) is chosen by type \( \theta \) only, is no better than a two-contract separation mechanism, \( \theta \in \{1, \beta\} \) in which type \( \theta \) only chooses \( \theta \) with probability 1. Since for each type, \( V_\theta(c_\theta) \) is the best value achieved in the original mechanism which is achieved in the separation mechanism with probability 1. Notice that \( c_\theta \) satisfies all the constraints from (63) to (67).

Case 2: A multiple contract mechanism in which one type always pool with the other (and the pooling type is \( \beta \), i.e., type 1 chooses contracts in \( \psi(\beta) \) with positive probability \( \psi(\beta) \) \leq 1. By Lemma 3, if \( c_i, c_j \in \psi(\beta) \), then \( a_i = a_j = a \). Consider mechanism \( \psi(\beta) \) in which both types pool. Pick a contract, \( \psi(\beta) \) such that \( E[\theta|c_p] \leq E[\theta|c(\beta)] \), where \( E[\theta|c(\beta)] = E[E[\theta|c_i] | c_i \in \psi(\beta)] \), is the conditional expectation of \( \theta \) on any contract \( c_i \in \psi(\beta) \) is chosen. We then modify \( c_p \) to \( c'_p \).
We let type \( \beta \) choose \( c'_p \) with probability 1 and type 1 choose it with probability \( \psi(\beta) \) instead. Thus, \( E[\theta|c'_p] = E[\theta|c(\beta)] \). We modify \( k_p \) so that \( a'_p = k_p r (1 - \alpha_p) \alpha_p E[\theta|c_p] = k_p r (1 - \alpha_p) \alpha_p E[\theta|c_p] = a \). 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 \( \frac{1}{2} k_i^2 + w_{p,0} = \frac{1}{2} k_j^2 + w'_{p,0} \). Thus \( c'_p = \{k'_p, \alpha_p, w'_{p,0}\} \). By construction, both types' payoff from choosing \( c'_p \) are the same as that of choosing a contract in \( \psi(\beta) \) in the original mechanism: For \( \theta = 1 \): If \( \psi(\beta) = 1 \), then the original mechanism is the same as this one contract, \( c'_p \) with both types pooling on it; If \( \psi(\beta) < 1 \), then there exists a contract \( c_1 \) chosen by type 1 with positive probability in the original mechanism. We now keep \( c_1 \) in the new mechanism and let type 1 chooses \( c_1 \) with probability \( 1 - \psi(\beta) \). The new mechanism has two contracts, \( c'_p \) and \( c_1 \) and type 1 is indifferent between the two contracts and type \( \beta \) strictly prefers \( c'_p \) and each type gets the same payoff as in the original mechanism. (The case where the pooling type is 1 is similarly proved.)

**Proposition 11** The principal’s expected payoff is maximized under pooling or separating mechanism.

**Proof:** We focus on \( \beta < \beta^* \) (since if \( \beta \geq \beta^* \), separation achieves the full information payoff). By Lemma 4, we consider only two contracts as in the separation case. However, a principal may select the contracts randomly, i.e., revealing information partially. Denote \( c_0 \) the contract a type \( \theta \) is more likely to choose. Now the mechanism is \( (c_0, \sigma_\theta) \) for \( \theta = 1, \beta \). \( \sigma_\theta \) is the probability that a type \( \theta \)
chooses $c_\theta$. The value of a type $\theta$ from choosing $c_\theta$ is $V^\theta_\theta$. Formally the contract design problem is

$$\max_{w_{e,0}, w_{\bar{e}}, k_\theta, \sigma_\theta} \sum_{\theta=1, \beta} p_\theta V^\theta_\theta$$

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

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

$V^\theta_\theta \geq V^\theta_{\hat{\theta}}$, for any $\hat{\theta} \neq \theta$

$(1 - \sigma_\theta)[V^\theta_\theta - V^\theta_{\hat{\theta}}] = 0$, for $\hat{\theta} \neq \theta$.

This problem adds two features to the separation case. First, the mechanism can be random. Second, there is a new constraint, (72), which is the complementary slackness constraint, i.e., if type $\theta$ is indifferent between the two contracts, randomization can occur. The previous program includes as particular cases full separation (i.e., $\sigma_\theta = 1$) and full pooling mechanisms (i.e., $c_\theta$ is the same for all $\theta$). There are three possible cases: (i) Only IC$_\beta$ in (71) binds; (ii) Only IC$_1$ in (71) binds; and (iii) Both ICs in (71) bind.

Case (i): If this case IC$_1$ can be ignored which is a contradiction with the fact that the full information allocation in subsection 2.2 satisfies all the constraints of the problem.

Case (ii): In this case, the type 1 uses a mixed strategy in choosing the two contracts, and the type $\beta$ 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 to relax IC$_1$ as much as possible as argued in the proof of proposition 2. Formally, the program is

$$\max_{w_{\beta,0}, w_{\beta}, k_\beta, \sigma_1} (1 - \pi)V^1_1 + \pi V^\beta_{\beta}$$

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

$V^1_1 \geq V^\beta_{\beta}$,

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

$k_\beta \geq 0$,

$0 \leq \sigma_1 \leq 1$.

By Bayes rule, $E[\theta|\hat{\theta} = \beta]e_\theta = 1+\frac{\pi}{\pi+1-\pi}(\beta-1)$. Let $E[\theta|\hat{\theta} = \beta] = v_\beta, v_\beta \in [(\beta)^{-1}, 1]$. We can change the control variable to from $\sigma_1$ to $v$ and consider the Lagrangian:

$L = (1-\pi)k^2 + \pi(rk_\beta(1-\alpha_\beta)\alpha_\beta rv^2 c - g(k_\beta)-w_{\beta,0})-\lambda(rk_\beta(1-\alpha_\beta)\alpha_\beta rv^2 c - g(k_\beta)-w_{\beta,0})k^2 + \mu w_{\beta,0} + v k_\beta \xi (\nu-1) .

By Kuhn-Tucker Theorem, we get the following FOCs with respect to $w_{\beta,0}, \alpha_\beta, k_\beta,$ and $v$:

$$\lambda - \pi + \mu = 0,$$

$$(\pi \beta - \lambda)(1 - 2\alpha_\beta) = 0,$$

$$(\pi \beta - \lambda)r(1 - \alpha_\beta)\alpha_\beta rv^2 c - (\pi - \lambda)k_\beta + v = 0,$$

$$(\pi \beta - \lambda)rk_\beta(1 - \alpha_\beta)\alpha_\beta rv^2 c - \xi = 0.$$

A similar argument as in the proof of Proposition 2 implies $\pi \beta - \lambda > 0$ which it turns implies $\xi > 0$, $v = 1$ and $\sigma_1 = 1$. Hence, the optimal solution is the separation case considered in the text.

Case (iii): In this case the solution would be inferior to the optimal pooling mechanism considered

37
section 3.2. Since both types are indifferent between the two contracts and the objective function only depends on the contract for type 1 (i.e., \( \max_{w_1,0,w_1,k_1,\sigma_\theta} \sum_{\theta=1,\beta} p_\theta V^1_\theta \)). Thus, w.l.o.g. \( E[\theta|c_1] \leq E[\theta] \) (i.e., if the type 1 contract is chosen, the agent’s posterior expectation of \( \theta \) is lowered) because there is always contract such that this holds for type 1. Now this contract is (weakly) inferior to the optimal pooling mechanism because otherwise we could define \((w_1,0,w_1,k_1)\) as the optimal pooling contract. In this case, the agent would exert a (weakly) higher effort in the optimal pooling mechanism 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 optimal pooling case too because (70) is satisfied and \( E[\theta|c_1] \leq E[\theta] \). Therefore, under \( c_1 \), the principal has a better payoff than under \((\tilde{w}^*,\tilde{k}^*)\), which is a contradiction. \( \blacksquare \)
Part 3: Proofs of Section 4

For this part we consider a general wage schedule, \( w(\theta, q) \) and define \( \alpha(\theta) = \frac{w(\theta, r_k(\theta)) - w(\theta, 0)}{r_k(\theta)} \).

**Proof of Proposition 5**

Denote \( V(\theta) = V^\theta_0 \), \( V(\theta, \hat{\theta}) = V^\theta_{\hat{\theta}} \), and \( k(\theta) = k^*_\theta \). (When there is no confusion, we drop the superscript \( s \) in the proof.) Proposition 5 follows directly from Lemma 5.

**Lemma 5** The conditions stated in Proposition 5 define the separating mechanism with the highest expected payoff to the principal, \( V(\theta) \).

**Proof.** In a separating mechanism, the principal reveals her information, i.e., \( \hat{\theta} = \theta \)

\[
\theta = \arg \max_\theta V(\theta, \hat{\theta}) = r^2 k(\hat{\theta})(1 - \alpha(\hat{\theta}))\alpha(\hat{\theta}) \frac{\hat{\theta}}{c} - \frac{1}{2} k(\hat{\theta})^2. \tag{83}
\]

Now we consider the case \( \alpha(\hat{\theta}) = \frac{1}{2} \) and \( k(1) = k^*_1 = \frac{r^2}{4c} \). The FOC yields:

\[
\frac{r^2}{4c} \theta^2 k(\theta) + \frac{r^2}{4c} k(\theta)\theta - k(\theta)k(\theta) = 0, \tag{84}
\]

\[
\dot{k}(\theta) = \frac{dk(\theta)}{d\theta}, \text{ with initial value } k(1) = k^*_1 = \frac{r^2}{4c}. \text{ For technical simplicity we solve } \theta(k) \text{ rather than } k(\theta) \text{ and we consider the unique positive solution of } \theta(k):
\]

\[
\theta = \sqrt{\frac{4c(2k(\theta)^3 + k^*_1)}{3k(\theta)r^2}}. \tag{85}
\]

The SOC of (83) requires that \( \frac{\partial V(\theta, \hat{\theta})}{\partial \theta^2} |_{\theta=\hat{\theta}} \leq 0 \). We denote \( V_i(\theta, \hat{\theta}), i = 1, 2 \), as the partial derivative with respect to the \( i \)’th argument and \( V_{ij}(\theta, \hat{\theta}) \) as the second order derivative. FOC implies \( V_2(\theta, \theta) = 0 \) for all \( \theta \) and therefore, \( \frac{dV(\theta, \theta)}{d\theta} = V_{12}(\theta, \hat{\theta}) + V_{22}(\theta, \theta) = 0 \). Thus if \( k(\theta) \theta \) is non-decreasing:

\[
V_{22}(\theta, \theta) = -V_{12}(\theta, \theta) = -\frac{r^2}{4c} \frac{dk(\theta)}{d\theta} \leq 0.
\]

Furthermore, if \( k(\theta) \theta \) is non-decreasing, we can show that the mechanism is sufficient for (83).

Suppose otherwise, \( V(\theta, \hat{\theta}) > V(\theta, \theta) \) for \( \theta \neq \hat{\theta} \) or \( \int_0^\theta V_2(\theta, x)dx > 0 \). Using FOC, this is equivalent to \( \int_0^\theta [V_2(\theta, x) - V_2(x, x)]dx > 0 \) or \( \int_\theta^0 r^2 V_{12}(y, x)dydx > 0 \). If \( \hat{\theta} > \theta \), then \( x > \theta \) and \( V_2(y, x) = \frac{r^2}{4c} \frac{dk(x)}{dx} \geq 0 \) which implies that the previous inequality does not hold. Similarly if \( \hat{\theta} < \theta \), then \( x < \theta \), and the inequality does not hold either.

Thus to show that the mechanism satisfies (83), we only need to show \( \frac{dk(x)}{dx} \geq 0 \). (By (84), \( \frac{dk(x)}{dx} = \frac{k(\theta)k(\theta)}{\frac{r^2}{4c} \theta} \), so \( \frac{dk(x)}{dx} \geq 0 \) iff \( k(\theta) \geq 0 \). But (84) implies \( \dot{k}(\theta)(\frac{r^2}{4c} \theta^2 - k(\theta)) = -\frac{r^2}{4c} k(\theta) \theta \) which implies \( \dot{k}(\theta) \geq 0 \) iff \( \frac{r^2}{4c} \theta^2 \leq k(\theta) \) (i.e., overinvestment). Thus, to show \( k(\theta) \geq 0 \) we only need to show that (85) has an unique root which is bigger than \( k^*_\theta = \frac{r^2}{4c} \theta^2 \). After some algebra, (85) is equivalent to:

\[
2k^3 + k^*_1 - 3k^2 k^*_\theta = 0 \quad \text{with} \quad k = k(\theta).
\]

Differentiating the left hand side (LHS) w.r.t. \( k \), we get \( 6k^2 - 6k^*_\theta k^* \), which is positive if \( k > k^*_\theta \). At \( k = k^*_\theta \), the LHS is \( k^*_1 - k^*_3 \leq 0 \), thus since the LHS is increasing there exists an unique root \( k \geq k^*_\theta \). (Notice that \( k = k^*_\theta \) only when \( \theta = 1 \).)

To show that this mechanism provides the highest payoff to the principal, assume on the contrary that \( \exists \) another separating mechanism for which \( V'(\theta_0) > V^*(\theta_0) \) for some \( \theta_0 \). Since \( V^*(1) \) is the full
information value, and \( V'(1) \leq V^*(1) \) then \( \theta_0 > 1 \). Let \( \theta_c = \sup \{ \theta | V'(\theta) \leq V^*(\theta) \} \) and \( \theta \in [1, \theta_0) \).

Since \( V'(\theta) - V^*(\theta) = 0 \) at \( \theta_c \), \( \exists \theta_1 \in [\theta_c, \theta_0) \) s.t. \( \frac{dV'(\theta)}{d\theta} \bigg|_{\theta_1} > \frac{dV^*(\theta)}{d\theta} \bigg|_{\theta_1} \). Furthermore \( V'(\theta_1) \geq V^*(\theta_1) \) by def. of \( \theta_1 \). That is:

\[
\begin{align*}
    r^2 k^s(\theta_1)(1 - \alpha^s(\theta_1))\alpha^s(\theta_1) &\leq r^2 k'(\theta_1)(1 - \alpha'(\theta_1))\alpha'(\theta_1) \\
    &\leq r^2 k'(\theta_1)(1 - \alpha'(\theta_1))\alpha'(\theta_1) \frac{\theta_1^2}{c} - \frac{1}{2} k'(\theta_1)^2 - w'(\theta_1, 0) \\
    &\leq r^2 k'(\theta_1)(1 - \alpha'(\theta_1))\alpha'(\theta_1) \frac{\theta_1^2}{c} - \frac{1}{2} k'(\theta_1)^2 - w'(\theta_1, 0)
\end{align*}
\]

The last inequality follows because \( \alpha^s(\theta_1) = \arg \max (1 - \alpha') \alpha \). Since \( k^s(\theta_1) \geq k^c(\theta_1) \) and \( \frac{1}{2} (r^2 k(1 - \alpha^c(\theta_1))\alpha^c(\theta_1)) \leq 0 \) for all \( k \geq k^c(\theta) \), then \( k^s(\theta_1) \geq k'(\theta_1) \). To compare \( \frac{dV(\theta)}{d\theta} \) at \( \theta = \theta_1 \), by the envelope theorem

\[
\frac{dV(\theta)}{d\theta} = \frac{dV(\theta, \theta)}{d\theta} = r^2 k(\theta)(1 - \alpha(\theta))\alpha(\theta) \frac{\theta}{c}
\]

which implies a contradiction:

\[
\begin{align*}
    \frac{dV'(\theta)}{d\theta} \bigg|_{\theta_1} &= r^2 k'(\theta_1)(1 - \alpha'(\theta_1))\alpha'(\theta_1) \frac{\theta_1}{c} \leq r^2 k'(\theta_1)(1 - \alpha^s(\theta_1))\alpha^s(\theta_1) \frac{\theta_1}{c} \leq \\
    &\leq r^2 k^s(\theta_1)(1 - \alpha^s(\theta_1))\alpha^s(\theta_1) \frac{\theta_1}{c} = \frac{dV^*(\theta)}{d\theta} \bigg|_{\theta_1}
\end{align*}
\]

Proof of Proposition 6

First we show that the optimal mechanism is a partition mechanism.

**Lemma 6** Without loss of generality the search for an optimal mechanism can be restricted to the set of partition mechanisms.

**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 type in \( \Theta(a) \) then it is separating. If there are multiple types, we know from Lemma 3 that only the boundary points (sup \( \Theta(a) \) and inf \( \Theta(a) \)) can choose more than one \( a \) and all types in between choose only \( a \). Thus we have pooling. In this context, pooling means the types choose the same \( a \) but the contracts can be different. We can show, however, that it is without loss of generality for the types that choose the same \( a \) pool together. To show that, notice that \( a = rk_c(1 - \alpha_c)E[c(\theta)] \) where \( c(\theta) \) is the contract chosen by \( \theta \) and let \( \alpha_c = \frac{w_c(k_c r) - w_c(0)}{r k_c} \). 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 one with least \( \frac{1}{2} k_c^2 + w_c(0) \) is the same. Furthermore, we can find an \( c_p \) such that \( a = (1 - \alpha_p)E[c(\theta)] \). That is, the pooling of all types choosing the same \( c_p \) can implement \( a \). To find \( c_p \) pick a contract with the same \( a \) such that \( E[c(\theta)] \leq E[c(\theta)] \) which always exists. If all types choosing \( a \) pool at the contract \( c \), then \( a' = rk(1 - \alpha_c)E[c(\theta)] \). But we can change \( \alpha_c \) to \( \alpha_p \) to lower \( a' \) to \( a \). Therefore w.l.o.g. we can consider one contract for the same \( a \).

Previous arguments imply that 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 \( \varphi^*_{\theta} \) be an interval between two separation intervals with \( \theta \in \varphi^*_{\theta} \). More precisely, for a given pooling interval \( [\bar{\theta}, \bar{\theta}] \), we can find a separation \( \varphi^*_\theta = \sup \{ \theta | \theta < \bar{\theta} \) and \( \theta \) is pooling \}, \( \bar{\theta} \).

Therefore the separation types are countable disjoint union of \( \varphi^*_{\theta} \). \( \blacksquare \)
We now simplify (27). At the boundary of an interval, $\bar{\theta}_i$ should be indifferent between choosing a contract in $\varphi_i$ and another in $\varphi_{i+1}$ in order to be incentive compatible, i.e.,

$$V(\bar{\theta}_i) = V(\underline{\theta}_{i+1}). \tag{87}$$

The left hand side is $\bar{\theta}_i$’s payoff if $\bar{\theta}_i$ chooses the contract in $\varphi_i$ and the right hand hand side is the payoff if $\bar{\theta}_i$ chooses the one in $\varphi_{i+1}$. Notice that (27) implies local incentive compatibility, i.e.,

$$V^0_{\theta_i} \geq V^0_{\theta_{i+1}}, \bar{\theta} \in \varphi(\theta), \tag{88}$$

that is, a type should not mimic another type in the same interval. But (87) and (88) imply (27) as the next lemma shows.

**Lemma 7** Condition (27) is equivalent to conditions (87) and (88).

**Proof.** Sufficiency is obvious. To prove necessity suppose (87) and (88) hold and (27) doesn’t. There are $\theta_1$ and $\theta_2$ such that $\theta_1 < \theta_2$ and $\theta_1$ is strictly better off mimicking $\theta_2$ (the case that $\theta_2$ mimicking $\theta_1$ is handled similarly). Hence (88) implies that $\varphi(\theta_1) \neq \varphi(\theta_2)$.

For ease of notation, for a contract $c_i$, let $\bar{c}_i \equiv \frac{1}{2} k_i^2 - w_i,0$. Consider two adjacent intervals, $\varphi_i$ and $\varphi_j$ with $\varphi_i = (\underline{\theta}_i, \bar{\theta}_i)$ between $\varphi(\theta_1)$ and $\varphi(\theta_2)$. We have $\theta_1 = \underline{\theta}_j$ so that $\varphi_i$ is below $\varphi_j$ (notice that $\varphi_j$ can be $\varphi(\theta_1)$ and $\varphi(\theta_2)$). For $\bar{\theta}_i$, (87) implies that $\bar{\theta}_i$ is indifferent between $\bar{c}_i$, which is chosen by a type in $\varphi_i$, and $\underline{c}_i$, which is chosen by a type in $\varphi_j$. That is $b_j - \underline{c}_i = (a_j - \underline{\theta}_i) \bar{\theta}_i \geq (a_j - \bar{\theta}_i) \bar{\theta}_j$ where the second inequality follows because the $\theta_1 \leq \theta_i$ and $a$ is non-decreasing by Lemma 3. By (88) $b_j - \underline{c}_i \geq (a_j - a_i) \bar{\theta}_j \geq (\bar{\theta}_j - a_j) \bar{\theta}_j$. Let $\bar{\theta}_j$ be the lower bound of $\varphi(\theta_2)$. By (88) $\bar{\theta}_j$ weakly prefers $\underline{c}_2$ to $c_2 = c(\theta_2)$, that is $b_2 - \underline{c}_2 \geq (a_2 - a) \bar{\theta}_2 \geq (\bar{\theta}_j - a_2) \bar{\theta}_j$. Similarly, on $\varphi(\theta_1)$ we have $\bar{\theta}_1 - \underline{c}_1 \geq (\underline{\theta}_1 - a_1) \bar{\theta}_1$ and summing up over all the inequalities (there could be countably infinite many of them) for intervals between $\theta_1$ and $\theta_2$, we get $b_2 - \bar{\theta}_1 \geq \theta_1(a_2 - a_1)$ which implies $\bar{\theta}_1$ prefers $c_1$ to $\underline{c}_2$ and hence a contradiction.

Therefore, we can replace (27) with (87) and (88) in the optimization problem and the next result says that we can relax (87). We use $t(\varphi(\theta)) \in \{p, s\}$ to denote an interval is pooling or separating.

**Lemma 8** The optimization problem yields the same solution if we relax (87) with $V(\bar{\theta}_i) \geq V(\underline{\theta}_{i+1})$. Further, in the optimal mechanism, $\alpha(\theta) = \frac{1}{2}$ and $w(\theta, 0) = 0$.

**Proof.** To prove the first part, assume that $k'(\cdot), \Psi', t'(\cdot)$ is a solution to the modified problem. We will show that $V(\bar{\theta}_i) \geq V(\underline{\theta}_{i+1})$ is binding at optimum. Let $\varphi'_i$ be an interval such that $V(\bar{\theta}_i) > V(\underline{\theta}_{i+1})$. If we show that all the types in $\varphi'_i$ are better off if we modify the solution by increasing $V(\underline{\theta}_{i+1})$ there arises a contradiction to the optimality of $k'(\cdot), \Psi', t'(\cdot)$. (This is because we improve the expected payoff of the firm with $V(\bar{\theta}_i) \geq V(\underline{\theta}_{i+1})$ for all $i$ and other constraints are satisfied.) If $\varphi'_{i+1}$ is separating, we consider $V(\underline{\theta}_{i+1}) = V(\bar{\theta}_i)$, $\alpha(\theta) = \frac{1}{2}$ and $w(\theta, 0) = 0$. Thus as in the proof of Proposition 5, the new separation improves the original separation since all types in $\varphi'_{i+1}$ are better off and the local IC (88) is satisfied. If $\varphi'_{i+1}$ is pooling, (88) is satisfied. Let $\alpha'_{i+1}, k'_{i+1}$ and $w'_{i+1}(0)$ as the contract for the types in $\varphi'_{i+1}$. First let $\alpha'_{i+1} \rightarrow \frac{1}{2}$ if $\alpha'_{i+1} \neq \frac{1}{2}$ so as to increase $(1 - \alpha'_{i+1}) \alpha_{i+1}$ and thus $V(\bar{\theta})$ on $\varphi'_{i+1}$. A similar argument shows that $w'_{i+1}(0) = 0$. If $\alpha'_{i+1} = \frac{1}{2}$ and $w'_{i+1}(0) = 0$, then $k'_{i+1}$ can be modified to increase $V(\bar{\theta}_i)$. This is always possible because $V(\bar{\theta}_i) > V(\underline{\theta}_{i+1}), k'_{i+1}$ does not maximize $V(\bar{\theta}_i)$. There are two possible changes to consider: an increase and a decrease in $k'_{i+1}$. An increase in $k'_{i+1}$ that increases $V(\bar{\theta}_i)$ implies that all $V(\bar{\theta})$ on $\varphi'_{i+1}$ increase because if $k'_{i+1}$ is an underinvestment for $\bar{\theta}_i$, it is also for all $\theta > \underline{\theta}_{i+1}$. In case of a decrease in $k'_{i+1}$ two pooling partitions on $\varphi'_{i+1}$, i.e., $\varphi_{i+1} = (\bar{\theta}_{i+1}, \theta_{i+1}]$ and $\varphi_{i+1}' = (\theta_{i+1}, \bar{\theta}_{i+1}]$ can
be created with the following features: a) the investment on $\varphi_{i+1}'$ is unchanged, i.e., $k_{i+1}' = k_{i+1}$. b) $\theta_{i+1}$ and the investment on $\varphi_{i+1}$, $k_{i+1}$, are chosen so that

$$V(\theta_{i+1} | \varphi_{i+1}) = V(\varphi_{i+1} | \varphi_{i+1}'),$$

and

$$V(\theta_{i+1} | \varphi_{i+1}) = V(\theta_{i+1} | \varphi_{i+1}').$$

(90)

For any $\theta_{i+1}$, there exists $k_{i+1}$ so that (89) holds. (Specifically, $\max_{k_{i+1}} \left( \frac{rk_{i+1}+\theta_{i+1}}{d} \right) E[ \theta | \theta_{i+1} | \theta_{i+1} | \varphi_{i+1}] - \frac{1}{2} k_{i+1}^2 > \frac{rk_{i+1} + \theta_{i+1}}{d} c(\theta_{i+1})$.) Notice that $k_{i+1} < k_{i+1}'$ because $E[ \theta | \theta_{i+1} | \theta_{i+1} | \varphi_{i+1}'] < E[ \theta | \theta_{i+1} | \theta_{i+1} | \varphi_{i+1}]$ and $V(\theta_{i+1} | \varphi_{i+1}', k_{i+1}) = V(\theta_{i+1} | \varphi_{i+1}', k_{i+1}')$ by assumption. As $\theta_{i+1} \rightarrow \theta_{i+1}$, $V(\theta_{i+1} | \varphi_{i+1}) \rightarrow V(\theta_{i+1} | \varphi_{i+1}')$ and $V(\theta_{i+1} | \varphi_{i+1}) \rightarrow V(\theta_{i+1} | \varphi_{i+1}')$ by continuity. Since $V(\theta_{i+1} | \varphi_{i+1}') < V(\theta_{i+1} | \varphi_{i+1})$ for $\theta_{i+1} \rightarrow \theta_{i+1}$, therefore, if (90) cannot hold for all $\theta_{i+1} \in \varphi_{i+1}$, (90) will hold as "\(" for all $\theta_{i+1} \in \varphi_{i+1}$. In this case:

$$V(\theta_{i+1} | \varphi_{i+1}, k_{i+1}) > V(\theta_{i+1} | \varphi_{i+1}, k_{i+1}').$$

(91)

The second inequality follows because $E[ \theta | \theta_{i+1} | \theta_{i+1} | \varphi_{i+1}] \geq E[ \theta | \theta_{i+1} | \varphi_{i+1}]$. Because $E[ \theta | \theta_{i+1} | \varphi_{i+1}] \geq E[ \theta | \theta_{i+1} | \varphi_{i+1}]$, $V(\theta_{i+1} | \varphi_{i+1}', k_{i+1}) > V(\theta_{i+1} | \varphi_{i+1}, k_{i+1}')$ which implies that $k_{i+1}$' improves all types on $\theta_{i+1} \in \varphi_{i+1}'.

If both (89) and (90) hold:

$$V(\theta | \varphi_{i+1}, k_{i+1}) > V(\theta | \varphi_{i+1}'', k_{i+1}').$$

(92)

for all $\theta \in \varphi_{i+1}$ by Lemma 3, $k_{i+1} < k_{i+1}'$ thus $a_{i+1} < a_{i+1}'$, and (90). As for types on $\varphi_{i+1}'$, Lemma 3, (90), and $a_{i+1} > a_{i+1}'$ (because $E[ \theta | \theta_{i+1} | \varphi_{i+1}] < E[ \theta | \theta_{i+1} | \varphi_{i+1}]$ and $k_{i+1}' = k_{i+1}'$) implies

$$V(\theta | \varphi_{i+1}') > V(\theta | \varphi_{i+1}'').$$

(93)

for $\theta \in \varphi_{i+1}'. Thus (92) and (93) imply that the modified mechanism is better for all types in $\varphi_{i+1}'$ which is a contradiction.

To prove the second part, assume that $\alpha_{i+1}' \neq \frac{1}{2}$. In this case, set $\alpha_{i+1} = \frac{1}{2}$ and increase $k_{i+1}'$ to $k_{i+1}$ so that $V(\theta_{i+1}) = V(\theta_{i+1} + k_{i+1}, a_{i+1} = \frac{1}{2})$. This increases $\alpha_{i+1}'$ and by Lemma 3 $V(\theta, \varphi_{i+1}', k_{i+1}, a_{i+1} = \frac{1}{2}) > V(\theta, \varphi_{i+1}', k_{i+1}, \alpha_{i+1})$ for $\theta \in \varphi_{i+1}'$. This increases firm value in the relaxed problem which is a contradiction. A similar argument shows if $\omega_{i+1}' > 0$, $k_{i+1}'$ can be increased to $k_{i+1}$ so that $V(\theta_{i+1}) = V(\theta_{i+1} + k_{i+1}, a_{i+1} = \frac{1}{2})$. This increases $\alpha_{i+1}'$ and by Lemma 3 $V(\theta, \varphi_{i+1}', k_{i+1}, a_{i+1} = \frac{1}{2}) > V(\theta, \varphi_{i+1}', k_{i+1}, \alpha_{i+1})$ for $\theta \in \varphi_{i+1}'$ which is a contradiction.

Part (1) of Proposition 6 follows directly from Lemma 8 and Part (2) follows from Lemmas 8 and 3. Thus the relaxed mechanism design problem is

$$\max \sum_{k_{i+1} \in \mathbb{R}} \int_{\varphi_{i+1}} V(\theta, k(\varphi)), t(\varphi(\theta))f(\theta) d\theta$$

s.t.

$$V(\theta) = \begin{cases} \frac{\sigma^2}{4} k(\theta) E[|\varphi(\theta)|] & \text{if } t(\varphi(\theta)) = p, \\
\frac{\sigma^2}{4} k(\theta) & \text{if } t(\varphi(\theta)) = s, \end{cases}$$

$$k(\theta) = \frac{\sigma^2}{4} k(\theta) \theta$$

(94)

$$V(\theta_{i+1}) \geq V(\theta_{i+1})$$

(95)

$$V(1) = \frac{\sigma^2}{4} k(1) E[|\varphi(1)|] - \frac{1}{2} k^2(1)$$

(96)

(97)
where $\dot{V}(\theta) = \frac{\partial}{\partial \theta} V(\theta, k(\varphi(\theta)), t(\varphi(\theta)))$ is the growth rate of $V$ and constraint $k(\theta) \geq 0$ is omitted. Expression (94) is derived from ICs by the envelope theorem and (95) is from the (84) of the firm’s IC as in the proof Proposition 5. Expression (97) says that the lowest type’s value is determined by $k(1)$ and the information revealed by choosing the contract $c(1)$.

Intuitively, the problem consists of finding $\Psi$, $t(\varphi(\theta))$ and $k(1)$. Once $\Psi$ and $t$ are fixed, for any given $k(1)$, $k(\theta)$ and $V(\theta)$ follow. That is, for any given interval with given initial $k(\theta_s)$, if $t = s$, (95) gives $k(\theta)$ and $V(\theta)$ on that interval. If $t = p$, then $k(\varphi(\theta)) = k(\theta_p)$ and $V(\theta)$ can be obtained in such interval. Finally, $k(\theta_{s+1})$ can be determined by (96) as an equality. The problem can thus be solved by choosing the best $k(1) \geq 0$.

Part (3) of Proposition 6 is proved by contradiction.

1) **Pooling at the top.**

We show that unless there is a pooling interval at the top with probability one, firm’s value can be improved. There are two cases.

Case (i): Separating at top.

Assume a separating mechanism on $(\theta_s, \bar{\beta})$ and to consider a mechanism with pooling on $(\theta_0, \bar{\beta})$ and that leaves untouched the rest of the proposed mechanism. Let $\theta_0$ be close to $\bar{\beta}$ so that $\theta_0 > \theta_s$.

Choose investment for the pooling interval $(\theta_0, \bar{\beta})$ such that type $\theta_0$ is indifferent between revealing itself and pooling with $(\theta_0, \bar{\beta})$, i.e., $V_P(\theta_0) = V^S(\theta_0)$. We denote $l = \beta - \theta_0$.

Step 1. On $\varphi(\theta) = (\theta_0, \bar{\beta})$, we solve for $k$ on $V(\theta) = \frac{r^2}{2}(k(\varphi(\theta))\theta E[\theta|\varphi(\theta)] - \frac{1}{2}k^2(\varphi(\theta))$ and get:

$$k = \frac{E[\theta|\varphi(\theta)] r^2 \theta + \sqrt{-32c^2 V(\theta) + r^4 \theta^2 E^2[\theta|\varphi(\theta)]}}{4c}$$

where the other root is ruled out since it implies underinvesting for $\theta_0$ but types in $\varphi(\theta)$ must overinvest to prevent mimicking from lower types. Substituting (98) into (94) gives:

$$\dot{V}(\theta) = \frac{r^2 E^2[\theta|\varphi(\theta)] + \sqrt{-32c^2 r^4 V(\theta) E^2[\theta|\varphi(\theta)] + r^8 \theta^2 E^2[\theta|\varphi(\theta)]}}{16c^2}$$

Step 2. Compare the payoff of the pooling case with full separation on $(\theta_0, \bar{\beta})$. Let $\theta_1 = \bar{\beta}$. If it is pooling in the general case, by Taylor Theorem on (99), $E[\theta|\theta \in (\theta_0, \bar{\beta})] = \theta_0 + \frac{1}{2} + O(l^2)$, and by using uniform distribution to approximate $f(\theta)$ we get

$$\dot{V}_P(\theta) = v_1(\theta_0) + \frac{1}{2} v_2(\theta_0) l + O(l^2), \theta \in [\theta_0, \theta_1]$$

where

$$v_1(\theta_0) = \frac{r^4 \theta_0^3 + \sqrt{-32c^2 r^4 V(\theta_0) \theta_0^2 + r^8 \theta_0^6}}{16c^2},$$

$$v_2(\theta_0) = \frac{-16c^2 r^4 V(\theta_0) \theta_0 + r^8 \theta_0^5 + 2 r^4 \theta_0^2 \sqrt{-32c^2 r^4 V(\theta_0) \theta_0^2 + r^8 \theta_0^6}}{8c^2 \sqrt{-32c^2 r^4 V(\theta_0) \theta_0^2 + r^8 \theta_0^6}},$$

$$l = \theta_1 - \theta_0.$$
where $O(\cdot)$ and $o(\cdot)$ are the usual big $O$ and small $o$ notation. Furthermore, because the remainder $O(l^2) = g(\theta)l^2$ where $\theta \in (\theta_0, \bar{\theta}_1]$ and $g$ is continuous in $\theta$, then $\bar{K}_p \leq g(\theta) \leq \underline{K}_p$ on $(\theta_0, \bar{\theta}]$.

Applying Taylor Th. on (99) and letting $V^s(\theta) = V(\theta_0) + v_1(\theta_0)(\theta - \theta_0) + O(\theta - \theta_0)^2$:

$$V^s(\theta) = v_1(\theta_0) + v_2(\theta_0)(\theta - \theta_0) + O(\theta - \theta_0)^2, \theta \in (\theta_0, \theta_1].$$

(101)

Similarly, $\bar{K}_s(\theta - \theta_0)^2 \geq O(\theta - \theta_0)^2 \geq \underline{K}_s(\theta - \theta_0)^2$ on $(\theta_0, \bar{\theta}]$. Therefore,

$$V^p(\theta) - V^s(\theta) \geq v_2(\theta_0) \int_{\theta_0}^{\theta} \frac{\theta_1 - \theta_0}{2} dx - v_2(\theta_0) \int_{\theta_0}^{\theta} (x - \theta_0) dx + \int_{\theta_0}^{\theta} [\underline{K}_p l^2 - \bar{K}_s(x - \theta_0)^2] dx$$

$$= \frac{v_2(\theta_0)}{2}(\theta_1 - \theta)(\theta - \theta_0) + \bar{K}l^3$$

(102)

Because $\bar{\theta}$ is separating and thus $V^s(\bar{\theta}) < \frac{\bar{\theta}^2}{\underline{K}_s}$ is the first best level, it follows that $v_2(\bar{\theta}) > 0$ from its definition. By continuity, $v_2(\theta_0) > 0$ for $l \to 0$:

$$\int_{\theta_0}^{\theta_1} [V^p(\theta) - V^s(\theta)] f(\theta) \geq \int_{\theta_0}^{\theta_1} \frac{v_2(\theta_0)}{2}(\theta_1 - \theta)(\theta - \theta_0) + \bar{K}l^3] f(\theta) d\theta \geq \frac{f_0 v_2(\theta_0)}{12} + f_0 \bar{K}l^4 > 0$$

for $l \to 0$. In the second inequality, $f(\theta_0) = f_0$ approximates the density function $f(\theta)$. (In a small neighborhood, the distribution function is approximately a uniform distribution.) From the last inequality and the fact that all types below $\theta_0$ are indifferent between the two mechanisms, we conclude that the modified mechanism dominates for $l$ small enough. Thus separation at the top can be ruled out. However, no separation at the top does not necessarily imply pooling at the top. Technically there can be infinite many intervals at the top, (some pooling, some separation) whose length goes to zero as it gets closer to $\bar{\theta}$. Since in such a case, there would not be a top interval with positive length so, strictly speaking, there would not be pooling at the top interval. Hence, to prove pooling at the top, we need to rule out infinite intervals clustering at top.

Case (ii): Clustering at the top.

Consider a set of infinite intervals in the neighborhood of $\bar{\theta}$. In this case, there always exists an arbitrarily small $\epsilon > 0$ so that $N \equiv (\bar{\theta} - \epsilon, \bar{\theta}) = \cup \varphi_i$. Let $\theta_0 = \bar{\theta} - \epsilon$. Similar to case (i), the idea is to show that a modified mechanism with pooling on $(\bar{\theta} - \epsilon, \bar{\theta})$ but everything else the same improves firm value. First we show that $V^p(\theta) - V^g(\theta) \geq 0$ on $(\bar{\theta} - \epsilon, \bar{\theta})$. Here the superscript $g$ stands for the general mechanism with clustering intervals. We then show that the expected firm value on $(\bar{\theta} - \epsilon, \bar{\theta})$ is strictly higher for the modified mechanism.

Fix a subinterval of $N$ of the general mechanism, $\varphi_i = (\bar{\theta}_i, \bar{\theta}_i]$. We have on $\varphi_i$

$$V^g(\theta) = V^g(\bar{\theta}_i) + \int_{\bar{\theta}_i}^{\theta} V'(x) dx \leq V^g(\bar{\theta}_i) + \int_{\bar{\theta}_i}^{\theta} \left[ v_1(\bar{\theta}_i) + v_2(\bar{\theta}_i) \frac{l_i^2}{2} \right] dx + K\bar{I}_i^3$$

where $l_i = \bar{\theta}_i - \bar{\theta}_i$ and $K > 0$. The inequality follows from (100) if it is pooling and from (102) if it is separating on $\varphi_i$. Applying Taylor Theorem at $\theta_0 = \bar{\theta} - \epsilon$ and $V^g(\theta) = V^g(\theta_0) + v_1(\theta_0)(\theta - \theta_0) + O(\theta - \theta_0)^2$ we get

$$v_1(\theta) = v_1(\theta, V^g(\theta)) = v_1(\theta_0) + v_2(\theta_0)(\theta - \theta_0) + \frac{1}{2} v_3(\theta')(\theta - \theta_0)^2, \theta' \in [\theta_0, \theta]$$

$$\leq v_1(\theta_0) + v_2(\theta_0)(\theta - \theta_0) + K_1(\theta - \theta_0)^2$$
because $v_3(\theta)$ is continuous on $N$, we can find a fixed $K_1$ such that $K_1 \geq \frac{1}{2}v_3(\theta)$ for $\theta \in N$. Thus,

$$V^g(\theta) \leq V^g(\theta_0) + \int_{\theta_0}^{\theta} \left[ v_1(\theta_0) + v_2(\theta_0)(\theta_0-\theta_0) + K_1(\theta_0-\theta_0)^2 + v_2(\theta_0) \frac{l_i}{2} \right] dx + Kl_i^3$$

$$= \int_{\theta_0}^{\theta} \left[ v_1(\theta_0) + v_2(\theta_0)(\theta_0-\theta_0) + K_1(\theta_0-\theta_0)^2 + v_2(\theta_0) \frac{l_i}{2} + v_2(\theta_0)(\theta_0-\theta_0) \right] \frac{l_i}{2} dx + Kl_i^3$$

$$\leq V^g(\theta_0) + v_1(\theta_0)(\theta-\theta_j) + v_2(\theta_0)(\theta_0-\theta_0) + K_1(\theta_0-\theta_0)^2 + K_2^2(\theta_0-\theta_0)^2 + Kl_i^3$$

$$\leq V^g(\theta_0) + v_1(\theta_0)(\theta-\theta_j) + v_2(\theta_0)(\theta_0-\theta_0) + K_3^2(\theta_0-\theta_0)^2. \quad (103)$$

The second inequality follows because we can find a $K_2$ such that $v_2(\theta) - v_2(\theta_0) \leq K_2(\theta - \theta_0)$ for all $\theta, \theta_0 \in N$ and the last inequality follows because $\theta_j - \theta_0 \leq \epsilon$ and $l_i \leq \epsilon$. At the end point $\theta_j$, (103) implies

$$V^g(\theta_j) - V^g(\theta_0) \leq v_1(\theta_0)(\theta_j - \theta_j) + v_2(\theta_0) \left[ \frac{\theta_j - \theta_0}{2} \right](\theta_j - \theta_0) + K_3^2 l_i^2$$

$$= v_1(\theta_0)(\theta_j - \theta_j) + \frac{v_2(\theta_0)}{2}[(\theta_j - \theta_0)^2 - (\theta_j - \theta_0)^2] + K_3^2 l_i^2. \quad (104)$$

Summing up both sides of (104) for all the intervals between $\theta_0$ and $\theta_j$,

$$\sum_{\varphi_j \leq \theta_j} V^g(\varphi_j) - V^g(\theta_0) \leq \sum_{\varphi_j \leq \theta_j} \left[ v_1(\theta_0)(\varphi_j - \theta_0) + \frac{v_2(\theta_0)}{2}[(\varphi_j - \theta_0)^2 - (\varphi_j - \theta_0)^2] + K_3^2 l_i^2 \right], \text{ or}$$

$$V^g(\theta) - V^g(\theta_0) \leq v_1(\theta_0)(\theta - \theta_0) + \frac{v_2(\theta_0)}{2}(\theta_0 - \theta_0)^2 + K_3^2(\theta_0 - \theta_0). \quad (105)$$

Substituting (105) into (103), we get

$$V^g(\theta) - V^g(\theta_0) \leq v_1(\theta_0)(\theta - \theta_0) + \frac{v_2(\theta_0)}{2}(\theta - \theta_0)^2 + v_1(\theta_0)(\theta - \theta_0) + \frac{v_2(\theta_0)}{2}(\theta - \theta_0)^2 + K_3^2(\theta_0 - \theta_0)$$

$$= v_1(\theta_0)(\theta - \theta_0) + \frac{v_2(\theta_0)}{2}[(\theta_0 - \theta_0)^2 - 2(\theta_0 - \theta_0)(\theta - \theta_0) + (\theta - \theta_0)(\theta - \theta_0)] + K_3^2(\theta_0 - \theta_0)$$

$$= v_1(\theta_0)(\theta - \theta_0) + \frac{v_2(\theta_0)}{2}[(\theta_0 + \theta_0 - \theta_0)^2 + (\theta_0 - \theta_0)(\theta - \theta_0)] + K_3^2(\theta_0 - \theta_0)$$

$$\leq v_1(\theta_0)(\theta - \theta_0) + \frac{v_2(\theta_0)}{2}[(\theta_0 + \theta_0 - \theta_0)^2 + (\theta_0 - \theta_0)(\theta - \theta_0)] + K_3^2(\theta_0 - \theta_0)$$

$$= v_1(\theta_0)(\theta - \theta_0) + \frac{v_2(\theta_0)}{2}(\theta_0 - \theta_0)(\theta_0 - \theta_0) + K_3^2(\theta_0 - \theta_0). \quad (106)$$

But in the modified mechanism in which it is pooling on $(\tilde{\beta} - \epsilon, \beta)$, (100) implies

$$V^p(\theta) - V^p(\theta_0) \geq v_1(\theta_0)(\theta - \theta_0) + \frac{v_2(\theta_0)}{2}(\theta_0 - \theta_0) + K_3^2(\theta - \theta_0). \quad (107)$$

Comparing (106) and (107) and notice the fact that $V^p(\theta_0) = V^g(\theta_0)$, we have

$$V^p(\theta) - V^g(\theta) \geq \frac{v_2(\theta_0)}{2}(\theta - \theta_0)(\theta - \theta_0) - M \epsilon^3 \quad (108)$$
where $M > |K_3| + |K_\rho|$. Integrating both sides of (108) on $N$,
\[
\int_{\theta_0}^{\bar{\theta}} [V^p(\theta) - V^q(\theta)] f(\theta) d\theta \geq \sum_j \int_{\theta_j}^{\bar{\theta}_j} \frac{v_2(\theta_0)}{2} (\theta_0 - \theta_0)(\epsilon - (\bar{\theta}_j - \theta_0)) f(\theta) d\theta - M \epsilon^3 (1 - F(\theta_0)),
\]
\[
\geq \int_{\theta_0}^{\bar{\theta}} \frac{v_2(\theta_0)}{2} (\theta_0 - \theta_0)(\epsilon - (\bar{\theta}_i - \theta_0)) f(\theta) d\theta - M \epsilon^3 (1 - F(\theta_0)),
\]
\[
\geq \frac{v_2(\theta_0)}{4} (\epsilon - (\bar{\theta}_i - \theta_0))(\bar{\theta}_i - \theta_0)^2 f_0 - M \epsilon^4 f_0
\]
\[
= f_0 \frac{v_2(\theta_0)}{4} (1 - \gamma) \gamma^2 \epsilon^3 - M \epsilon^4.
\]  
(109)

Here $\gamma \equiv (\bar{\theta}_i - \theta_0)/\epsilon$ and notice that $0 \leq \gamma \leq 1$. The second inequality follows because the integrand is positive and $\bar{\theta}_i > \theta_0$. (The third inequality follows by using $f_0$ to approximate $f(\theta)$ and ignoring higher order terms.) The last term is positive if $(1 - \gamma) \gamma^2$ is bounded away from zero as $\epsilon$ goes to zero. That is, if there is a sequence of $\epsilon_n \downarrow 0$ and $\bar{\theta}_i^n$ such that $(1 - \gamma_n) \gamma_n^2 \geq C > 0$, $\exists \epsilon_n$ small enough so that $\int_{\theta_0}^{\bar{\theta}_i^n} [V^p(\theta) - V^q(\theta_0)] f(\theta) d\theta > 0$. That is, the modified mechanism is strictly better.

If such a sequence of $\epsilon_n$ does not exist then $(1 - \gamma_n) \gamma_n^2 \to 0$ for any sequence of $\epsilon_n \downarrow 0$. Let $\epsilon_m \downarrow 0$. We will construct a subsequence $\epsilon_{n \downarrow}$. We start with an arbitrary $\epsilon_0 > 0$. For all the intervals $\varphi_i$ in $(\beta - \epsilon_0, \beta]$, we pick the one with the maximum length $\varphi_i^* = (\bar{\theta}_i, \bar{\theta}_i)$, for such an interval, $\varphi_i^*$ has the maximum length. Let $\gamma_n \equiv (\bar{\theta}_i - \theta_0)/\epsilon_n$. Therefore we have $(1 - \gamma_n) (\gamma_n^1)^2 \to 0$. There are two cases. If there is a subsequence $\gamma_j \to 0$, then for $\epsilon_j$ small enough we can find an interval in $(\theta_0, \epsilon_j/3, \theta_0 + \epsilon_j/2)$. We denote $\theta_{0,j} = \theta_j^1$. For such an interval, $\gamma_j \equiv (\bar{\theta}_j - \theta_{0,j})/\epsilon_j$ and $(1 - \gamma_j) \gamma_j^2 \geq \frac{1}{2}(1 - \epsilon_j)^2 > 0$. Thus, (109) implies that for $\epsilon_j$ small enough, we have modified mechanism strictly better than the general mechanism. If we cannot find a subsequence $\gamma_j \to 0$, then $(1 - \gamma_n) (\gamma_n^1)^2 \to 0$ implies that $\gamma_n \to 1$. That is
\[
\lim_{n \to \infty} \frac{l_{1,n}}{\epsilon_n} = 1,
\]
(110)

where $l_{1,n} \equiv (\bar{\theta}_i - \theta_i)/\epsilon_n$ is the length of the first interval. Notice that it is pooling on $\varphi_1^n$ for $\epsilon_n$ small. (otherwise we divide $\varphi_1^n$ into two intervals so as to construct $(1 - \gamma_n) (\gamma_n^1)^2 \to C > 0$.)

We show (110) is not optimal by showing that a full pooling is better on $(\beta - \epsilon_n, \beta]$ for $\epsilon_n$ small enough. Notice that $l_{1,n}$ is the dominant partition in that $\epsilon_n - l_{1,n}$ goes to zero. On $\varphi_1^n$, since it is pooling, substituting $E[\theta | \varphi_1^n] = \theta_0 + l_{1,n}/2 + O(l_{1,n}^2)$ into (99) gives
\[
V(\theta, l_{1,n}) = \frac{1}{16c^2} [r^2 \theta_{0,n} (\theta_{0,n} + l_{1,n}/2 + O(l_{1,n}^2))^2
\]
\[
+ \sqrt{-32c^2 r^4 V(\theta_{0,n}) [\theta_{0,n} + l_{1,n}/2 + O(l_{1,n}^2)]^2 + r^8 \theta_{0,n}^2 (\theta_{0,n} + l_{1,n}/2 + O(l_{1,n})^4].
\]
46
Similarly, for the full pooling, $l_n = \epsilon_n$

$$
\dot{V}(\theta, l_n) = \frac{1}{16c^2} [r^2 \theta_{0,n}(\theta_{0,n} + \frac{l_n}{2} + O(l_n^2)]^2 \\
+ \sqrt{-32c^2 r^4 V(\theta_{0,n})(\theta_{0,n} + \frac{l_n}{2} + O(l_n^2))]^2 + r^8 \theta_{0,n} (\theta_{0,n} + \frac{l_n}{2} + O(l_n^2))}.
$$

Thus by Taylor Theorem

$$
\dot{V}(\theta, l_n) - \dot{V}(\theta, l_{1,n}) = \frac{v_2(\theta_{0,n})}{2}(l_n - l_{1,n}) + o(l_n - l_{1,n}). \quad (111)
$$

As a convention, $o(l_n - l_{1,n})$ denotes a term such that $\lim_{l_n \to 0} \frac{o(l_n - l_{1,n})}{l_n} = 0$ as $l_n - l_{1,n}$ goes to 0. Therefore by (111), on $\varphi_{1,0}$

$$
\int_{\theta_{0,n}}^{\varphi_{1,0}} [V^p - V^q] f(\theta) \theta \Theta = \int_{\theta_{0,n}}^{\varphi_{1,0}} \left[ \frac{v_2(\theta_{0,n})}{2} (l_n - l_{1,n}) + o(l_n - l_{1,n}) \right](\theta - \theta_{0,n}) f(\theta) d\theta \\
\geq \frac{v_2(\theta_{0,n})}{2} f_{0,n} \frac{l_{1,n}^2}{2} (l_n - l_{1,n}) + o (l_n - l_{1,n}) l_{1,n}^2
$$

where $f_{0,n} = f(\theta_{0,n})$. Therefore on $(\varphi_{1,0}, \bar{\beta})$ by (108),

$$
\int_{\varphi_{1,0}}^{\bar{\beta}} [V^p - V^q] f(\theta) \Theta \geq \int_{\varphi_{1,0}}^{\bar{\beta}} -M_3 \epsilon_{n}^3 f(\theta) \Theta \geq -M \epsilon_{n}^3 (l_n - l_{1,n}) \bar{T} = 0(l_n - l_{1,n}) l_{1,n}^2
$$

where $\bar{T}$ is the maximum of $f(\theta)$ on $(\bar{\beta} - \epsilon_1, \bar{\beta})$. The last equality follows because (110). Therefore,

$$
\int_{\theta_{0,n}}^{\varphi_{1,0}} [V^p - V^q] f(\theta) \Theta \geq \frac{v_2(\theta_{0,n})}{2} f_{0,n} \frac{l_{1,n}^2}{2} (l_n - l_{1,n}) + o (l_n - l_{1,n}) l_{1,n}^2 > 0
$$

for $\epsilon_n = l_n$ close to zero which is inconsistent with clustering at the top.

2) Pooling at the bottom

Assume that non-pooling at the bottom maximizes firm value. There can be two cases:

Case (i) Separation at the bottom.

Let $\theta_1 \in [\theta_0 = 1, \theta_2]$ and assume it is separation. Hence IC of $\theta_0$ implies

$$
V^s(\theta_0) \geq \frac{r^2}{4c} \theta_0 \theta_1 k_1 - \frac{1}{2} k_1^2 \quad (112)
$$

where $k_0$ is the investment of $\theta_0$ and $k_1$ is that of $\theta_1$. Thus

$$
V^s(\theta_1) = \frac{r^2}{4c} \theta_1^2 k_1 - \frac{1}{2} k_1^2 \leq V^s(\theta_0) + \frac{r^2}{4c} (\theta_1 - \theta_0) \theta_1 k_1. \quad (113)
$$

The inequality follows from (112). Notice that $V^s(\theta_0) = V^s(\theta_0)$, since otherwise firm value would improve for types in $[\theta_0, \theta_2]$ by setting $V^s(\theta_0) = V^s(\theta_0)$. (This argument is similar to the proof of Proposition 5.)
If there is pooling on \([\theta_0, \theta_s]\), let \(\theta^p = E[\theta|\theta \in [\theta_0, \theta_s]]\) and \(k_p = \arg \max_k \frac{r^2}{4c} \theta^2 p^2 k - \frac{1}{2}k^2\) be the investment level for the interval. Thus for \(\theta_p\), \(V^p(\theta_p) = V^*(\theta_p)\) is the first best value. By (94),

\[
V^p(\theta_1) = V^p(\theta_p) + \frac{r^2}{4c} \theta_p k_p (\theta_1 - \theta_p) = V^*(\theta_0) + \frac{r^2}{4c} \theta_p k_p (\theta_1 - \theta_p) + o(\theta_p - \theta_0). \tag{114}
\]

Notice that: \(V^p(\theta_p) = V^*(\theta_0) + \frac{r^2}{4c} 2\theta_0 k^*_p (\theta_p - \theta_0) + o(\theta_p - \theta_0)\) where the second equality follows from applying the envelope theorem to \(V^*(\theta_0)\) and thus \(\frac{dV^*}{d\theta} = \frac{r^2}{4c} 2\theta_0 k^*_p\) and by Taylor Theorem. Combining (114) and (113) yields

\[
V^p(\theta_1) - V^*(\theta_1) \geq V^*(\theta_0) - V^*(\theta_0) + \frac{r^2}{4c} \theta_0 k^*_0 [2(\theta_p - \theta_0) + (\theta_1 - \theta_p)] - \frac{r^2}{4c} \theta_1 k_1 (\theta_1 - \theta_0) + o(\theta_p - \theta_0)
\]

\[
\geq \frac{r^2}{4c} \theta_0 k^*_0 [2(\theta_p - \theta_0) + (\theta_1 - \theta_p) - (\theta_1 - \theta_0)] + o(\theta_p - \theta_0)
\]

\[
= \frac{r^2}{4c} \theta_0 k^*_0 (\theta_p - \theta_0) + o(\theta_p - \theta_0)
\]

\[
= \frac{r^2}{4c} \theta_0 k^*_0 \left(\frac{\theta_s - \theta_0}{2}\right) + o(\theta_p - \theta_0) \tag{115}
\]

The second inequality follows because \(V^*(\theta_0) - V^*(\theta_0) = 0\) and \(\theta_1 k_1 - \theta_0 k^*_0 \to 0\) as \(\theta_s \to \theta_0\) by continuity of \(k(\theta)\). The last equality follows because \(\theta_p = \theta_0 + \frac{\theta_s - \theta_0}{2} + o(\theta_s - \theta_0)\). (115) implies that \(V^p(\theta_1) - V^*(\theta_1) > 0\) for \(\theta_s\) close to \(\theta_0\). We thus improve the solution to the relaxed problem as in Lemma 8 which is a contradiction.

Case (ii) Clustering at the bottom.

In this case, \(\theta_0\) is fully revealed so \(V^g(\theta_0) = V^*(\theta_0)\). If \(V^g(\theta_0) < V^*(\theta_0)\) by continuity there is an upper end point of an interval \(\theta_1 > \theta_0\) such that \(V^g(\theta) \leq V^*(\theta)\) for any \(\theta \in [\theta_0, \theta_1]\). Hence, by considering a pooling mechanism on \([\theta_0, \theta_1]\) with investment \(k^*_0\) (without affecting the rest of the mechanism) we improve firm value: \(V^p(\theta_0) \geq V^*(\theta_1) > V^g(\theta_0)\). (This is so since \(V^p(\theta) > V^p(\theta_0)\) for \(\theta \in [\theta_0, \theta_1]\).) Thus the solution to the relaxed problem as stated in Lemma 8.

We can find \(N \equiv (1, 1 + \varepsilon) = \bigcup_{i \in N} \varphi_i\) for arbitrarily small \(\varepsilon > 0\). Let \(\theta_0 = 1\) and \(\theta_s = 1 + \varepsilon\). Similar to case (i), it is possible to modify the original mechanism \(g\) such that it is pooling on \([\theta_0, \theta_s]\) and keep everything else the same.

Fix \(\theta_1 \in N\) and \(\theta'_1 = E[\theta|\varphi(\theta_1)]\). If it is separation on \(\varphi(\theta_1)\), \(\theta'_1 = \theta_1\). Denote \(k_1\) the investment level for \(\theta_1\). Similar argument in (113) yields

\[
V^g(\theta_1) = \frac{r^2}{4c} \theta_1 \theta'_1 k_1 1 \leq V^g(\theta_0) + \frac{r^2}{4c} \theta'_1 k_1 (\theta_1 - \theta_0) = V^*(\theta_0) + \frac{r^2}{4c} \theta'_1 k_1 (\theta_1 - \theta_0). \tag{116}
\]

The pooling on \([\theta_0, \theta_s]\) is as in case (i), i.e., \(V^p(\theta_1)\) is given by (114). Comparing (116) with (114):

\[
V^p(\theta_1) - V^g(\theta_1) \geq \frac{r^2}{4c} 2\theta_0 k^*_p (\theta_p - \theta_0) + \frac{r^2}{4c} \theta_p k_p (\theta_1 - \theta_p) - \frac{r^2}{4c} \theta'_1 k_1 (\theta_1 - \theta_0) + o(\theta_p - \theta_0)
\]

\[
\geq \frac{r^2}{4c} \theta_0 k^*_0 [2(\theta_p - \theta_0) + (\theta_1 - \theta_p) - (\theta_1 - \theta_0)] + o(\theta_p - \theta_0) \tag{117}
\]
The second inequality follows because $\theta^*_1 \rightarrow \theta_0$ and $k_1 \rightarrow k^*_0$ as $\theta_s \rightarrow \theta_0$. To show that $k_1 \rightarrow k^*_0$ notice that $\theta_1$’s IC implies

$$\frac{r^2}{4c} \theta_1 \theta_1' k_1 - \frac{1}{2} k_1^2 \geq \frac{r^2}{4c} \theta_0 k^*_0 - \frac{1}{2} k^*_0^2$$

which is

$$\frac{r^2}{4c} \theta_1 \theta_1' (k^*_0 - k_1) + \frac{k^*_0 + k_1}{2} (k^*_0 - k_1) = (k_1 - k^*_0)(\frac{k^*_0 + k_1}{2} - \frac{r^2}{4c} \theta_1 \theta_1')$$

$$= (k_1 - k^*_0)(\frac{k^*_0 + k_1}{2} - k^*_0) - (k_1 - k^*_0)(\theta_1 \theta_1' - \theta_0^2) + \frac{(\theta_1 \theta_1' - \theta_0^2)^2}{2} - \frac{(\theta_1 \theta_1' - \theta_0^2)^2}{2}$$

$$= \frac{1}{2} [k_1 - k^*_0 - (\theta_1 \theta_1' - \theta_0^2)]^2 - \frac{(\theta_1 \theta_1' - \theta_0^2)^2}{2}. \quad (118)$$

The second equality follows from $k^*_0 = \frac{r^2}{4c} \theta_0^2$. Hence (118) is equivalent to

$$k_1 - k^*_0 \leq \sqrt{\frac{r^2}{2c} \theta_1 \theta_1' (k^*_0 - k_1)} + (\theta_1 \theta_1' - \theta_0^2)^2 + (\theta_1 \theta_1' - \theta_0^2)^2$$

whose right hand side goes to zero as $\theta_s \rightarrow \theta_0$. Thus $k_1 - k^*_0 \rightarrow 0$ as $\theta_s \rightarrow \theta_0$.

Since (117) is equivalent to (115), the rest of the proof is as in case (i).

For part (4), the separating case follows from the proof of Proposition 5 that shows overinvestment with probability 1. The proof of the pooling case is by contradiction and starts by assuming

$$\tilde{k}_{\varphi_i}^* < \arg \max_k \{E[V^\theta_{\tilde{\theta_i}} | \varphi_i^{p^*}] \}. \quad (119)$$

To ease notation, let $k_i \equiv \tilde{k}_{\varphi_i}^*$ and $k^*_i \equiv \arg \max_k \{E[V^\theta_{\tilde{\theta_i}} | \varphi_i^{p^*}] \}$ and consider the following cases:

First, if increasing $k_i$ lowers $V(\tilde{\theta}_i)$ then increasing $k_i$ by a small amount, $\Delta k$, to $k_i'$ would increase the expected firm value on $\varphi_i^{p^*}$. In particular, $V(\tilde{\theta}_i)$ increases because (119) implies that $k_i$ is under-investing for type $\theta = E[\tilde{\theta}_i | \varphi_i^{p^*}]$ and thus for $\tilde{\theta}_i$ too. Hence $k_i'$ satisfies the relaxed problem as in stated in Lemma 8 and would improve the firm’s expected value which is a contradiction.

Second, if increasing $k_i$ increases $V(\tilde{\theta}_i)$, then, since $V(\tilde{\theta}_i)$ is quadratic in $k_i$, $\exists k_i' > k_i$ such that $V(\tilde{\theta}_i)$ remains unchanged (i.e., $k_i'$ and $k_i$ are the two roots). Therefore all types on $\varphi_i^{p^*}$ are better off under $k_i'$ by Lemma 3. Since $k_i'$ satisfies the relaxed problem, this is a contradiction.■
References


