ESSAYS IN INFORMATION ECONOMICS

By

John Andrew Withers

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Economics — Doctor of Philosophy

2018

ABSTRACT

ESSAYS IN INFORMATION ECONOMICS

By

John Andrew Withers

The first chapter of this dissertation studies a repeated interaction between a regulator and a regulated firm. In each period, the firm completes a project for the regulator, and the regulator observes the project's cost. The firm's intrinsic cost level is a component of the project's cost. Thus, the regulator gathers information about the firm's intrinsic cost level by observing the project's cost. This information is valuable to the regulator; the more she knows about the firm's intrinsic cost level, which is fixed over time, the more efficient is the outcome of their interaction in each period.

An important feature of the interaction is that the project's cost is stochastic; that is, the firm has imperfect control over the project's cost. The firm determines the expected project cost by choosing its effort, but a noise term determines the cost realization. The first chapter demonstrates that the regulator's first period contract choice determines how much she learns about the firm's intrinsic cost level. The main contribution of the first chapter is to show that, given a reasonable assumption about the distribution of noise, the low cost firm's first period effort is lower than his second period effort. This result aligns with anecdotal, experimental and empirical evidence of the ratchet effect.

The second chapter examines an interaction that is similar to the first chapter, with one important difference: the agent's productivity, which is akin to his intrinsic cost level in the first chapter, is positively correlated over time, rather than fixed. Unlike the standard ratchet effect literature, the low productivity agent has an incentive to reveal information to the principal. If the high ability agent is not too much more productive than the low ability

agent, or if the high productivity agent is sufficiently likely ex-ante, the optimal first period contract restricts what the principal learns about the agent's first period type.

The third chapter considers a two period contracting problem between one principal, one agent, and an outside labor market. In the first period, the principal hires the agent to exert unverifiable effort on a project that may either succeed or fail. Effort can be high or low. In the second period, the labor market makes the agent a wage offer if the project is successful. The principal has the opportunity to match the outside offer, or let the agent leave the firm. When the agent leaves the firm, the principal incurs a cost of replacing the agent.

The agent is "self motivated." That is, the expected value of the outside offer is high enough that the agent prefers high effort to low effort in the absence of an incentive wage. When the cost of replacing the agent exceeds a certain threshold, the principal prefers low effort to high effort, even though the agent is self motivated.

ACKNOWLEDGMENTS

I am grateful for the countless hours that my adviser, Thomas Jeitschko, spent discussing this dissertation with me over the past four years. Without his encouragement, guidance and mentorship, it would not have been possible for me to finish. I would also like to thank my other committee members, Mike Conlin, Arijit Mukherjee, and Adam Candeub, for their helpful comments and advice.

I would like to thank Lori Jean Nichols for her support and expert administrative help during my time as a graduate student at Michigan State University. I would also like to thank Todd Elder for helping guide me through the daunting task of finding a job.

Most importantly, I would like to thank parents, my sister and my fiancé for their unwavering love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Dynamic Regulation with Stochastic Costs: Signal Dampening, Experimentation and the Ratchet Effect

1.1 Introduction

In regulated industries, firms and regulators have long-term relationships with one another. The rules and procedures that govern these relationships are revised over time. When the regulator cannot commit at the outset of the relationship to how these rules and procedures will be updated in the future, the ratchet effect arises.

In repeated principal-agent interactions, the ratchet effect describes the agent's response to the principal's inability to commit to long term contracts. The principal learns about the agent's ability, or the economic environment, by observing his performance. The principal then adjusts the agent's compensation in the future based on what she learns from this observation. The more the principal learns about the agent, the more rent she is able to extract. To obscure the principal's learning process, the agent restricts his performance, or reduces his effort. This allows the agent to avoid more stringent incentives in the future.

Take, for example, a regulated monopoly that provides electricity to consumers. Periodically, the regulator will undertake a rate case to evaluate whether current electricity prices offer the utility a fair return on capital. During the rate case, the regulator observes the utility's operating expenses, along with other measures such as the firm's rate base (capital), taxes and depreciation expenses. Based off these measures, the regulator determines the revenue that the firm needs to earn to recoup operating expenses and make a fair return for their investors. This revenue target in turn determines the prices that the utility can charge consumers.

During this process, the regulator learns about the firm's efficiency by observing the firm's operating expenses. The regulator expects that a firm with high operating expenses in the current rate cycle will have high operating expenses again in the next rate cycle, and is thus more willing to give a generous reimbursement. Therefore, the firm has little incentive reduce operating costs, since a better performance today implies a less generous revenue requirement in the next rate cycle.

Some of the earliest anecdotal evidence of the ratchet effect comes from studies of piece rate factory workers (see Matthewson (1931), Roy (1952), Montgomery (1979) and Clawson (1980)). Matthewson (1931) documented that piece-rate workers understood that a good performance today ultimately made them worse off in the long run. To see this, suppose a worker produces more units of output in the current period than in the previous pay period. Since the worker is paid per unit, the worker earns more in the current period than in the previous period. Workers learned, however, that the factory manager's response to this improved performance was to reduce the worker's piece rate. Thus, the worker had to keep producing a high output just to earn as much take home pay as they did before revealing favorable information about their productive ability. In response to this behavior by factory managers, Matthewson documented that workers "never worked at anything like full capacity." Berliner (1957) documented that factory managers in the Soviet Union responded similarly to incentive systems based on output targets.

The anecdotal evidence discussed above suggests that agents restrict their performance (i.e., reduce effort) when the principal bases their future compensation on information that she gathers about them. Recent empirical evidence supports this notion. Macartney (2016) adapts the theoretical model of Weitzman (1980) to examine if teacher value-added schemes induce dynamic effort distortions among teachers in North Carolina. Teachers in a given school receive a bonus in the current year if the school-wide average on a standardized test is above a pre-specified target. The key feature of these schemes is that the target score is a function of the school's average standardized test score in the previous year. Clearly, the higher is the school's average test score this year, the more difficult it will be for teachers to exceed next year's target and receive a bonus. Macartney exploits differences in grade composition across schools to show that teachers respond to the value-added schemes by reducing their effort on improving their students standardized test scores.

In the kind of repeated interactions described by Matthewson (1931) and Macartney (2016), agents with high ability have the strongest incentive to reduce effort in the present to maintain information rents in the future. Charness, Kuhn and Villeval (2011) use an experimental design to study the effects of labor market competition on the ratchet effect. As a baseline case, they examine a two-period relationship between one firm and one worker. In this baseline case, roughly 60 percent of the experimental subjects who are designated as having high ability reduce their effort in the first period so that they can maintain a second period information rent. In a related experimental paper, Cardella and Depew (2018) study the impact of evaluating performance at the individual versus group level on the ratchet effect. The authors find that workers suppress effort when evaluated individually.

In most theoretical models of the ratchet effect, the good agent's effort does not evolve as one would expect based on the anecdotal, empirical, and experimental evidence discussed above. For example, Laffont and Tirole (1987) examine a two-period interaction between a regulator and a regulated firm in which the firm completes a project for the regulator. The observable outcome is the project's cost. The project cost depends on the firm's intrinsic cost level, which is the firm's private information. The regulator cannot commit, in the first period, to the second period incentive scheme.

In this setting, the low-cost firm exerts the first best level of effort in the first and second period unless he places a large enough weight on the second period contract. One reason the low-cost firm's effort in Laffont and Tirole (1987) does not evolve in a manner that fits with received evidence is because the firm is assumed to have perfect control over the observable outcome. That is, the only way for the low cost agent to hide his private information is to mimic (pool with) the high cost firm.

Contrast this with the case in which the agent does not have perfect control over the observable outcome (i.e., the relationship between the agent's actions and the project's outcome is stochastic). In the framework of Laffont and Tirole (1987), this can be achieved by assuming project costs depend on an additive, zero-mean noise term. Laffont and Tirole (1986) and Laffont and Tirole (1993) show that an additive, zero-mean noise term has no impact on incentives in a static setting.[1]

In a dynamic setting, however, noise plays the crucial role of slowing the principal's learning process. Jeitschko, Mirman and Salgueiro (2002) and Jeitschko and Mirman (2002)

[1] This assumes that both the firm and the regulator are risk neutral.

study two-period interactions in which an agent produces output for a principal. Output in each period depends on the agent's effort, his inherent productivity, and a zero-mean noise term. The agent's productivity is his private information and can take one of two values. In each period, the agent's compensation depends only on observed output.

In this setting, the agent's effort choice determines the expected output level. In equilibrium, the agent chooses his effort so that his expected output is equal to an output target proposed by the principal. Therefore, the principal's choice of equilibrium output targets determines the distribution of output for each type of agent in each period. For this reason, the principal's second period beliefs are a function of her first period contract choice.

Jeitschko et al. (2002) and Jeitschko and Mirman (2002) show that two opposing incentives determine the first period output targets. First, the principal can design the first period contract to increase what she learns about the agent's private information. By doing so, she increases her expected second period payoff. Second, the principal can design the first period contract to decrease what she learns about the agent's private information. By doing so, she decreases the first period transfer to the high productivity agent.

This paper examines a two-period model of regulation. In each period, a firm completes a project for a regulator. The observable outcome is the project's cost. The project's cost is stochastic, and the principal uses the cost observation to update her beliefs about the firm's type. We show that when the noisy component of the project's cost follows a general distribution, the low-cost agent has his effort increased over time. Therefore, we present a theoretical model whose predictions match with anecdotal, empirical and experimental evidence of the ratchet effect.[2]

[2] Jeitschko et al. (2002) assume the noise follows a uniform distribution and show that it is optimal for the high ability agent to exert less than the first best effort in the first period, which in turn implies that his first period effort is lower than his second period effort. Jeitschko and Mirman (2002) examine a similar setting in which the distribution of noise is general and are unable to determine how the high ability agent's

This paper is related to two strands of dynamic principal-agent literature. First, this paper is related to theoretical models of the ratchet effect. The ratchet effect has most famously been studied in the context of regulation and procurement (Freixas, Guesnerie and Tirole (1985), Laffont and Tirole (1987), Laffont and Tirole (1988) and Laffont and Tirole (1993)). It has also been studied in settings such as piece-rate incentive contracts (Gibbons (1987)), optimal income taxation (Dillen and Lundholm (1996)), and government corruption (Choi and Thum (2003)). These papers differ from the current paper in that the agent is assumed to have perfect control over the observable outcome.

This paper is also related to a growing dynamic mechanism design literature. Athey and Segal (2013) and Pavan, Segal and Toikka (2014) derive efficient and revenue maximizing dynamic mechanisms, respectively, when the principal can commit to future mechanisms and the agent's private information changes over time (for a survey of dynamic mechanism design when the principal can commit to future incentive schemes, see Bergemann and Valimaki (2017)). Because the principal is assumed to commit to future mechanisms, the ratchet effect problem does not arise.

The dynamic mechanism design literature most closely related to this paper studies dynamic mechanisms in which the principal has limited commitment power. First, Skreta (2015) studies a two period model in which a seller cannot commit not to re-sell an indivisible good if the first period mechanism fails to allocate the good to one of several buyers. Deb and Said (2015) study a sequential screening problem that builds off of Courty and Li (2000). The seller can commit in the first period to the terms of consumption of a good in the second period, but cannot commit to the selling mechanism offered in the second period. The principal in both Skreta (2015) and Deb and Said (2015) is concerned with maximizing

effort evolves over time.

revenue, while the principal in our paper maximizes welfare. Additionally, consumption only occurs once in each paper; in either the first or second period in Skreta (2015), and at the end of the second period in Deb and Said (2015). In our paper, the agent completes a task for the principal in each period. The principal gathers information about the agent from the outcome of the first period project, and uses this information to increase the efficiency of the second period interaction.

Finally, Gerardi and Maestri (2017) study an infinitely repeated principal-agent interaction. The principal is uninformed about the agent's private cost characteristic, which may be high or low. The agent produces a good of observable and verifiable quality for the principal. Depending on the principal's prior beliefs and the discount factor, the principal learns the agent's type immediately, over time, or never at all. Because Gerardi and Maestri (2017) study a pure adverse selection setting, there are no direct comparisons between our paper and theirs about how the low cost agent's effort evolves over time.

1.2 Model

Consider a two period interaction between a welfare-maximizing regulator (she) and a regulated firm (he). In each period, the regulator offers the firm a contract to complete a project that has gross-benefit S. In return for completing the project each period, the regulator reimburses the firm for the project's cost, c_t, and pays the firm an additional transfer, $t_t(c_t)$. The additional transfer is a function of the project's realized cost in each period, and incentivizes cost-reducing effort. The project's cost in each period depends on the firm's intrinsic cost parameter, β, its unobservable effort, e_t, and a homoskedastic, zero mean noise term,

ε_t:

$$c_t = \beta - e_t + \varepsilon_t, \quad t = 1, 2. \tag{1.1}$$

The random variable ε_t is assumed to be distributed over the entire real line according to the distribution function $G(\varepsilon)$ with associated density $g(\varepsilon)$. The density satisfies the monotone likelihood ratio property. While the full support assumption is analytically convenient, it raises two issues that bear mention.

The first issue is that the low cost firm's effort from mimicking the high cost type may be negative in the second period. This occurs when the first period cost realization is sufficiently low. A common assumption in static models is that the regulator's prior belief that the firm has low costs is small enough that this situation does not arise. However, in this dynamic-stochastic setting, the regulator's second period beliefs are endogenous, and depend on the first period cost realization. Thus, the analysis allows for negative effort levels. Second, the full support assumption implies that negative cost realizations are possible. While unrealistic, the possibility of negative costs does not affect the results of this paper.

It is important to note that ε_t is unobservable both ex-ante and ex-post. Thus, while the regulator is able to observe total cost c_t in each period, she cannot determine the individual impacts of the firm's type, its effort, and noise. This captures the intuition that the firm does not have perfect control over the project's cost. The firm affects the distribution of costs by exerting effort, but the project's cost depend on factors outside of the firm's control. Another interpretation of noise is that of an "accounting error." Given the complexity of accounting rules, and constraints on her time, the regulator may not able to perfectly discern which costs should and shouldn't be reimbursed after observing the firm's income statement or other supporting documents.

The firm's type can be either $\underline{\beta}$ or $\bar{\beta}$, with $0 < \underline{\beta} < \bar{\beta}$, and remains constant over the course of the interaction. Throughout, type $\underline{\beta}$ is referred to as the "low cost type" or "low cost firm," and type $\bar{\beta}$ as the "high cost type" or "high cost firm." The firm's type is its private information; the regulator's prior belief that the firm is the low cost type is given by ρ. The firm experiences a disutility of effort that can be expressed in monetary terms by

$$\psi(e_t) = \begin{cases} \frac{\gamma}{2} e_t^2, & e_t > 0, \\ 0, & e_t \leq 0, \end{cases} \tag{1.2}$$

where $\gamma > 0$. Thus, the firm's per period utility is given by

$$U_t = t_t(c_t) - \psi(e_t). \tag{1.3}$$

Although project costs are stochastic, the firm's effort is not; in each period, the firm chooses his effort before the realization of ε_t.

The regulator's objective in each period is to maximize expected welfare, which is the sum of taxpayer surplus and the firm's utility. In each period, welfare is given by

$$W_t = S - (1 + \lambda)\big(c_t + t_t(c_t)\big) + U_t. \tag{1.4}$$

Taxpayers enjoy benefit S from the project, compensate the firm for its costs c_t, and pay out the incentive fee $t_t(c_t)$. Since the cost reimbursement and incentive transfer are raised via distortionary taxation, one dollar paid to the firm costs taxpayers $\$(1 + \lambda)$, where $\lambda > 0$ denotes the shadow cost of public funds.

The solution concept used is that of a perfect Bayesian equilibrium. In each period, the regulator designs an incentive scheme to maximize expected welfare. The incentive scheme depends on the regulator's beliefs about the firm's type. In the first period, the regulator considers the impacts of the first period contract on expected second period welfare.

At the beginning of the second period, the regulator observes the first period project cost, and updates her beliefs about the firm's type using Bayes' rule. Contracts are short term; thus, when designing the second period contract, the regulator cannot commit to ignore any information she learns about the firm's type from observing the realized first period project cost.

The firm chooses whether to participate or not in each period. If the firm chooses to participate, he chooses his effort to maximize his expected utility given the transfer designed by the regulator. In the first period, he considers the impact that his actions have on the regulator's second period beliefs, and thus his expected second period payoffs.

In the analysis to follow, the regulator's problem in each period is to maximize expected welfare by choosing a cost target for each type of firm. These targets serve two purposes. First, whatever cost the firm decides to target determines the firm's effort. To see this, recall that effort is chosen before the realization of ε_t. Thus, the firm simply chooses its effort such that its expected cost, $E[c_t] = \beta - e_t$, is equal to its chosen cost target.

Second, for a given type of firm, the cost target serves as the mean of the distribution of project costs in each period. Since the incentive transfer is a function of project costs, the expected transfer in each period depends on the cost target. Thus, at the beginning of each period the regulator chooses cost targets that, in expectation, form an incentive feasible menu.

Framing the regulator's problem as a choice of cost target for each type of firm is without loss of generality as long as there exists an incentive transfer, based solely on realized costs, that satisfies the three following properties in expectation. First, the high cost firm's expected utility from targeting $\bar{c}_t$ must be equal to his outside option of zero. Second, the low cost firm's expected utility from targeting $\underline{c}_t$ must be equal to his expected utility from targeting $\bar{c}_t$. Third, the firm's expected utility from targeting $c_t \notin \{\underline{c}_t, \bar{c}_t\}$ is lower than his expected utility from targeting either $\underline{c}_t$ or $\bar{c}_t$.

When these three properties are satisfied, the high cost firm's participation constraint and the low cost firm's incentive constraint are satisfied in expectation in each period. Further, neither firm has an incentive to target a cost level other than the cost target designed for him by the regulator. The paper proceeds by assuming that there exists a transfer based on observed costs, $t_t(c_t)$, such that the expected transfer, $E[t_t(c_t)]$, satisfies the three aforementioned properties.

Caillaud, Guesnerie and Rey (1992), Picard (1987) and Melumad and Reichelstein (1989) study the existence of such reward schedules when the agent's type space is continuous. When the agent's type may only take on two values, there are fewer constraints placed on the reward schedule. However, the lower envelope of the high and low cost agent's indifference curves is kinked, which implies that it may not be possible to implement the high cost firm's exact cost target. However, one can implement a cost target that is arbitrarily close (see Jeitschko and Mirman (2002)).

Throughout the paper, the focus is on deriving an equilibrium that is "separating in actions." Because cost observations are noisy, and this uncertainty is not resolved ex-post, the regulator is not able to determine with certainty the firm's type by on observing the cost realization. That is, even when the first period contract is designed in a way that the low cost firm and high cost firm target distinct cost levels, the regulator does not have full information about the firm's type in the second period. Thus, the equilibrium is separating in actions when the regulator designs distinct targets for each type of firm, and each type of firm targets the expected cost designed for for him by the regulator. This means in period $t = 1, 2$, the low cost firm targets $\underline{c}_t$, and the high cost firm targets $\overline{c}_t$.

1.3 Second period

Since the model is solved using backward induction, the analysis begins with the second period. Suppose that the first period contract is separating in actions. At the beginning of the second period, the regulator observes the first period cost realization and updates her beliefs about the firm's type using Bayes' rule. Therefore, her second period belief that the firm is the low cost type is given by

$$\rho_2 := \frac{\rho g(c_1 - \underline{c}_1)}{\rho g(c_1 - \underline{c}_1) + (1 - \rho) g(c_1 - \overline{c}_1)}. \tag{1.5}$$

Consider the numerator of (1.5). The regulator's prior belief that the firm has low costs is given by ρ. In the first period, the low cost firm targets $\underline{c}_1$; when the firm targets $\underline{c}_1$, the first period cost realization is $c_1 = \underline{c}_1 + \varepsilon_1$. Since $g(\varepsilon_1)$ represents the density of noise in the first period, $g(c_1 - \underline{c}_1)$ is the probability density of first period costs when the agent targets $\underline{c}_1$. Thus, $g(c_1 - \underline{c}_1)$ gives the value of the probability density function when the cost realization is c_1 and the agent targets $\underline{c}_1$.

Similarly, the probability density of costs when the agent targets $\overline{c}_1$ is given by $g(c_1 - \overline{c}_1)$. Since noise has full support on the real line, both $g(c_1 - \overline{c}_1)$ and $g(c_1 - \underline{c}_1)$ are strictly positive on the entire real line. Thus, the principal never believes to be fully informed about the agent's type in the second period. That is, because of the full support assumption,

$\rho_2 \in (0,\, 1)$.

With beliefs given in (1.5), the regulator's problem is to choose expected costs $\underline{c}_2$ and $\bar{c}_2$ to maximize expected welfare, subject to incentive and participation constraints (which are derived below):

$$\max_{\underline{c}_2, \bar{c}_2} \quad \rho_2 \int_{\mathbb{R}} \left[S - (1+\lambda)\big(c_2 + t_2(c_2)\big) + t_2(c_2) - \frac{\gamma}{2}(\underline{\beta} - \underline{c}_2)^2\right] g(c_2 - \underline{c}_2)dc_2$$
$$+ (1-\rho_2) \int_{\mathbb{R}} \left[S - (1+\lambda)\big(c_2 + t_2(c_2)\big) + t_2(c_2) - \frac{\gamma}{2}(\bar{\beta} - \bar{c}_2)^2\right] g(c_2 - \bar{c}_2)dc_2. \quad (1.6)$$

Because the second period game is static, and both the regulator and the firm are risk neutral, zero-mean noise has no impact on incentives. Thus, the binding constraints on the regulator's problem are the low cost type's incentive compatibility constraint and the high cost firm's participation constraint.[3]

First, consider the low cost type's incentive compatibility constraint. The optimal second period cost targets make the low cost firm's expected utility from targeting $\underline{c}_2$ equal to his expected utility from targeting $\bar{c}_2$. When the low cost firm targets $\underline{c}_2$, he chooses his effort in the second period such that $\underline{e}_2 = \underline{\beta} - \underline{c}_2$, and thus his private cost of effort is equal to $\frac{\gamma}{2}\left(\underline{\beta} - \underline{c}_2\right)^2$.

When the low cost firm chooses his effort in this manner, it is easy to see that

$$E[c_2] = E\left[\underline{\beta} - \underline{\beta} + \underline{c}_2 + \varepsilon_2\right] = \underline{c}_2. \quad (1.7)$$

Therefore, the second period project cost can be written as $c_2 = \underline{c}_2 + \varepsilon_2$, which implies that the density of second period costs is given by $g(c_2 - \underline{c}_2)$. Therefore, the low cost firm's

[3]The low cost firm's incentive constraint depends on whether the low cost type's effort from mimicking the high cost type is positive or negative. This issue is addressed shortly.

expected second period utility from targeting $\underline{c}_2$ is given by

$$E\left[\underline{U}_2 \mid \underline{c}_2\right] := \int_{\mathbb{R}} \left[t_2(c_2) - \frac{\gamma}{2}(\underline{\beta} - \underline{c}_2)^2\right] g(c_2 - \underline{c}_2) dc_2 = \underline{t}_2 - \frac{\gamma}{2}\left(\underline{\beta} - \underline{c}_2\right)^2, \tag{1.8}$$

where $\underline{t}_2 := \int_{\mathbb{R}} t_2(c_2) \cdot g(c_2 - \underline{c}_2) dc_2$.

Similarly, when the low cost type targets $\bar{c}_2$, his effort is given by $\bar{e}_2 - \Delta\beta = \underline{\beta} - \bar{c}_2$, and the density of second period costs is given by $g(c_2 - \bar{c}_2)$. Thus, his expected utility from targeting $\bar{c}_2$ is

$$E\left[\underline{U}_2 \mid \bar{c}_2\right] := \int_{\mathbb{R}} \left[t_2(c_2) - \frac{\gamma}{2}(\underline{\beta} - \bar{c}_2)^2\right] g(c_2 - \bar{c}_2) dc_2 = \bar{t}_2 - \frac{\gamma}{2}\left(\underline{\beta} - \bar{c}_2\right)^2, \tag{1.9}$$

where $\bar{t}_2 := \int_{\mathbb{R}} t_2(c_2) \cdot g(c_2 - \bar{c}_2) dc_2$. The low cost firm's incentive compatibility constraint makes him indifferent, in expectation, between targeting $\underline{c}_2$ and $\bar{c}_2$:

$$E\left[\underline{U}_2 \mid \underline{c}_2\right] = E\left[\underline{U}_2 \mid \bar{c}_2\right] \quad \Longrightarrow \quad \underline{t}_2 - \frac{\gamma}{2}\left(\underline{\beta} - \underline{c}_2\right)^2 = \bar{t}_2 - \frac{\gamma}{2}\left(\underline{\beta} - \bar{c}_2\right)^2. \tag{1.10}$$

The second period game is designed to extract all expected rent form the high cost type. When the high cost type targets $\bar{c}_2$, his cost of effort is $\bar{e}_2 = \bar{\beta} - \bar{c}_2$, and the density of expected costs is given by $g(c_2 - \bar{c}_2)$. Thus, the high cost type's expected second period rent is given by

$$E\left[\bar{U}_2 \mid \bar{c}_2\right] := \int_{\mathbb{R}} \left[t_2(c_2) - \frac{\gamma}{2}(\bar{\beta} - \bar{c}_2)^2\right] g(c_2 - \bar{c}_2) dc_2 = \bar{t}_2 - \frac{\gamma}{2}\left(\bar{\beta} - \bar{c}_2\right)^2. \tag{1.11}$$

Therefore, the high cost type's participation constraint is given by

$$E\left[\overline{U}_2 \,|\, \bar{c}_2\right] = 0 \quad \Longrightarrow \quad \bar{t}_2 - \frac{\gamma}{2}\left(\bar{\beta} - \bar{c}_2\right)^2 = 0. \tag{1.12}$$

Simplifying the objective function in (1.6) and using (1.10) and (1.12) to substitute for the expected transfers leaves the following unconstrained problem:

$$\begin{aligned} \max_{\underline{c}_2, \bar{c}_2} \quad & S - \rho_2 \left[(1+\lambda)\left(\underline{c}_2 + \frac{\gamma}{2}(\underline{\beta} - \underline{c}_2)^2\right) + \lambda\left(\frac{\gamma}{2}(\bar{\beta} - \bar{c}_2)^2 - \frac{\gamma}{2}(\underline{\beta} - \bar{c}_2)^2\right)\right] \\ & - (1-\rho_2)(1+\lambda)\left(\bar{c}_2 + \frac{\gamma}{2}(\bar{\beta} - \bar{c}_2)^2\right), \end{aligned} \tag{1.13}$$

where $\frac{\gamma}{2}(\bar{\beta} - \bar{c}_2)^2 - \frac{\gamma}{2}(\underline{\beta} - \bar{c}_2)^2$ is the low cost firm's expected information rent.

The first order conditions of this problem imply the following equilibrium efforts and cost targets:

$$\underline{e}_2 = \underline{\beta} - \underline{c}_2 = \frac{1}{\gamma}, \tag{1.14}$$

and

$$\bar{e}_2 = \bar{\beta} - \bar{c}_2 = \frac{1}{\gamma} - \frac{\rho_2}{1-\rho_2}\frac{\lambda}{1+\lambda}\Delta\beta. \tag{1.15}$$

Thus, the low cost type exerts the first best effort in the second period, and the high cost type's effort is distorted away from the first best according to the principal's second period beliefs. Notice that the effort levels given in (1.14) and (1.15) correspond to the standard static game in which beliefs are given by ρ_2. This illustrates than in a static setting, additive noise has no impact on incentives when the regulator and firm are risk neutral.

One concern in this model is that the low cost firm's effort from mimicking the high cost firm,

$$\bar{e}_2 - \Delta\beta = \underline{\beta} - \bar{c}_2 = \frac{1}{\gamma} - \frac{1 + \lambda - \rho_2}{(1 - \rho_2)(1 + \lambda)}\Delta\beta, \tag{1.16}$$

can be less than zero for values of ρ_2 close to one. "Negative effort" captures any measures taken to increase the project's cost. To understand why the low cost type might have to increase the project's cost to mimic the high cost type, recall that the expected cost for the high cost type is equal to its type minus its cost reducing effort. When the first period cost observation is low, this leads the regulator to believe that she is very likely to be contracting with the low cost type in the second period. In response, she reduces the effort of the high cost type in order to extract rent from the low cost type. When this effort is small enough (i.e. when ρ_2 is close to one), $\bar{c}_2 = \bar{\beta} - \bar{e}_2 > \underline{\beta}$.

This possibility is usually assumed away in static models. However, as ε has full support on the real line, it must be considered in this setting. Since g satisfies the monotone likelihood ratio property, the principal's posterior belief that the firm has low costs is monotone decreasing in first period cost realizations. Therefore, there exists a unique value of ρ_2, defined

$$\rho_2^0 := \rho_2(c_1^0) = \frac{(1 + \lambda)(1 - \gamma\Delta\beta)}{1 + \lambda - \gamma\Delta\beta} < 1, \tag{1.17}$$

such that for every $c_1 \leq c_1^0$, the low cost type's effort from mimicking the high cost type is negative.

Since the firm cannot experience a dis-utility from negative effort (that is, $\psi(e_t) = 0$ when $e_t \leq 0$), the low cost type's second period incentive compatibility constraint is written

$$\underline{t}_2 - \frac{\gamma}{2}(\underline{\beta} - \underline{c}_2)^2 = \bar{t}_2. \tag{1.18}$$

The high cost firm's participation constraint remains unchanged. Together, this implies that the regulator's unconstrained problem when $c_1 \leq c_1^0$ is given by

$$\max_{\underline{c}_2, \bar{c}_2} \quad S - \rho_2 \left[(1+\lambda) \left(\underline{c}_2 + \frac{\gamma}{2}(\underline{\beta} - \underline{c}_2)^2 \right) + \lambda \frac{\gamma}{2} \left(\bar{\beta} - \bar{c}_2 \right)^2 \right] \\ - (1-\rho_2)(1+\lambda) \left(\bar{c}_2 + \frac{\gamma}{2}(\bar{\beta} - \bar{c}_2)^2 \right), \tag{1.19}$$

where the low cost firm's expected information rent is now given by $\frac{\gamma}{2} \left(\bar{\beta} - \bar{c}_2 \right)^2$.

The first order condition for this problem with respect to $\bar{c}_2$ implies the following equilibrium effort for the high cost type (the low cost type still exerts the first best effort):

$$\bar{e}_2^0 = \bar{\beta} - \bar{c}_2 = \frac{1}{\gamma} \frac{(1-\rho_2)(1+\lambda)}{1+\lambda-\rho_2}. \tag{1.20}$$

The following proposition summarizes the second period game:

Proposition 1.1. *When $c_1 > c_1^0$, the regulator's problem is given by (1.13), while for $c_1 \leq c_1^0$, the regulator's problem is given by (1.19). The first order conditions of (1.13) and (1.19) with respect to $\underline{c}_2$ and $\bar{c}_2$ imply that the low cost firm's equilibrium expected rent is given by*

$$\underline{U}_2(\rho_2) = \begin{cases} \frac{\gamma}{2}(\bar{e}_2)^2 - \frac{\gamma}{2}(\bar{e}_2 - \Delta\beta)^2 =: \underline{u}_2, & \text{if} \quad c_1 > c_1^0, \\ \frac{\gamma}{2}(\bar{e}_2^0)^2 =: \underline{u}_2^0, & \text{if} \quad c_1 \leq c_1^0, \end{cases} \tag{1.21}$$

where $\bar{e}_2$ is given in (1.15), $\bar{e}_2 - \Delta\beta$ in (1.16), and $\bar{e}_2^0$ in (1.20). Similarly, equilibrium *expected second period welfare is given by*

$$W_2(\rho_2) = \begin{cases} S - \rho_2 \left[(1+\lambda)\left(\underline{\beta} - \frac{1}{2\gamma}\right) + \lambda \underline{u}_2\right] - (1-\rho_2)(1+\lambda)\left(\bar{\beta} - \bar{e}_2 + \frac{\gamma}{2}(\bar{e}_2)^2\right) =: w_2, \\ S - \rho_2 \left[(1+\lambda)\left(\underline{\beta} - \frac{1}{2\gamma}\right) + \lambda \underline{u}_2^0\right] - (1-\rho_2)(1+\lambda)\left(\bar{\beta} - \bar{e}_2^0 + \frac{\gamma}{2}(\bar{e}_2^0)^2\right) =: w_2^0, \end{cases} \tag{1.22}$$

when c_1 is greater than c_1^0 and less than c_1^0, respectively.

Regardless of the size of c_1, the second period game exhibits the classic rent extraction/efficiency trade-off present in static adverse selection models:

$$\frac{d\underline{U}_2(\rho_2)}{d\rho_2} = \begin{cases} \frac{d\underline{u}_2}{d\bar{e}_2}\frac{d\bar{e}_2}{d\rho_2} = \frac{-1}{(1-\rho_2)^2}\frac{\lambda}{1+\lambda}\gamma\Delta\beta^2 < 0, & \text{if} \quad c_1 > c_1^0, \\ \frac{d\underline{u}_2^0}{d\bar{e}_2^0}\frac{d\bar{e}_2^0}{d\rho_2} = \frac{-\lambda(1+\lambda)^2}{\gamma}\frac{1-\rho_2}{(1+\lambda-\rho_2)^3} < 0, & \text{if} \quad c_1 \leq c_1^0. \end{cases} \tag{1.23}$$

This is an important consideration for the regulator in the first period, since ρ_2 is a function of $\underline{c}_1$ and $\bar{c}_1$.

To see how second period beliefs, and thus second period welfare, depend on the first period contract, consider $\tilde{c}_1 = \underline{c}_1 + x$, for some fixed value x. From (1.5), the closer together are $\underline{c}_1$ and $\bar{c}_1$, the closer together are the values of $\underline{g}(\tilde{c}_1)$ and $\bar{g}(\tilde{c}_1)$. The closer together are $\underline{g}(\tilde{c}_1)$ and $\bar{g}(\tilde{c}_1)$, the closer ρ_2 is to the prior, ρ; indeed, if $\underline{c}_1 = \bar{c}_1$, then $\underline{g}(\tilde{c}_1) = \bar{g}(\tilde{c}_1)$ for all x, and the posterior is equal to the prior. Conversely, the further apart are $\underline{c}_1$ and $\bar{c}_1$, the smaller is $\bar{g}(\tilde{c}_1)$ relative to $\underline{g}(\tilde{c}_1)$, and the closer the posterior is to one.

Thus, the distance between first period cost targets directly influences how much the regulator updates her prior, given a first period cost realization. The further apart are the first period cost targets, the more accurate are the regulator's second period beliefs; the more accurate are the regulator's second period beliefs, the closer second period welfare is to the first-best. However, this information comes at a cost. Since the low cost firm's second period rent is decreasing in ρ_2, spreading the cost targets apart decreases (in expectation) the low cost firm's rent from targeting $\underline{c}_1$, and increases his rent from targeting $\bar{c}_1$ in the first period. This increases the low cost type's first period transfer. Thus, the regulator faces a tradeoff between increasing the expected second period welfare *or* preserving the low cost firm's expected second period rent.

1.4 First period

The second period beliefs, ρ_2, serve as the link between the first and second period contracts. When choosing the first period cost targets, the regulator considers not only the impact that they have on first period welfare, but what impact they have on expected second period welfare as well. The regulator's first period problem is to maximize the expectation of first and (discounted) second period welfare, subject to incentive compatibility and participation constraints, which are derived below:

$$\begin{aligned} \max_{\underline{c}_1, \bar{c}_1} \qquad & S - \rho \int_{\mathbb{R}} \left[(1+\lambda)\left(c_1 + t_1(c_1)\right) + t_1(c_1) - \frac{\gamma}{2}\left(\underline{\beta} - \underline{c}_1\right)^2 \right] g(c_1 - \underline{c}_1) dc_1 \\ & - (1-\rho) \int_{\mathbb{R}} \left[(1+\lambda)\left(c_1 + t_1(c_1)\right) + t_1(c_1) - \frac{\gamma}{2}\left(\bar{\beta} - \bar{c}_1\right)^2 \right] g(c_1 - \bar{c}_1) \\ & + \delta E[W_2(\rho_2)], \end{aligned} \tag{1.24}$$

where $W_2(\rho_2)$ is given in (1.22), and

$$E[W_2(\rho_2)] = \int_{\mathbb{R}} W_2(\rho_2) \left[\rho g(c_1 - \underline{c}_1) + (1-\rho) g(c_1 - \bar{c}_1) \right] dc_1. \tag{1.25}$$

A well known issue in dynamic games is that the first period payment to the low cost firm may be so large that the high cost type's incentive compatibility constraint binds (the so-called "take the money and run" strategy). For now, consider the low cost firm's incentive compatibility constraint and the high cost firm's participation constraint.[4] The low cost firm's incentive constraint requires that his expected utility from targeting $\underline{c}_1$ equal his expected utility from targeting $\bar{c}_1$. That is,

$$\begin{aligned} E[\underline{U}_1 | \, \underline{c}_1] &:= \int_{\mathbb{R}} \left[t_1(c_1) - \frac{\gamma}{2}(\underline{\beta} - \underline{c}_1)^2 + \delta \underline{U}_2(\rho_2) \right] \underline{g} dc_1 \\ &= \int_{\mathbb{R}} \left[t_1(c_1) - \frac{\gamma}{2}(\underline{\beta} - \bar{c}_1)^2 + \delta \underline{U}_2(\rho_2) \right] \bar{g} dc_1 =: E[\underline{U}_2 | \, \bar{c}_2], \end{aligned} \tag{1.26}$$

where $\underline{g} := g(c_1 - \underline{c}_1)$ and $\bar{g} := g(c_1 - \bar{c}_1)$. The left hand side of (1.26) is the low cost firm's expected utility when he targets $\underline{c}_1$ in the first period. He exerts effort $\underline{e}_1 = \underline{\beta} - \underline{c}_1$, and receives an expected first period transfer and expected second period rent, where expectations are taken over the real line according to the density $\underline{g}$. If the low cost firm instead chooses to target $\bar{c}_1$, he experiences a disutility from effort $\bar{e}_1 - \Delta\beta = \underline{\beta} - \bar{c}_1$, and receives an expected first period transfer and expected second period rent. These expectations are taken according to the density $\bar{g}$.

From the perspective of the high cost firm, the first period game is essentially static since the second period game extracts all the rent from the high cost type. Therefore, the high cost firm's participation constraint requires that his expected first period utility from targeting $\bar{c}_1$ be equal to his outside option of zero:

$$E[\bar{U}_1 | \, \bar{c}_1] := \int_{\mathbb{R}} \left[t_1(c_1) - \frac{\gamma}{2} \left(\bar{\beta} - \bar{c}_1 \right)^2 \right] \bar{g} dc_1 = 0. \tag{1.27}$$

[4] In sufficiently noisy environments, the high cost firm's incentive constraint is slack. See Appendix A.

By defining $\underline{t}_1$ and $\bar{t}_1$ analogously to $\underline{t}_2$ and $\bar{t}_2$, one can simplify (1.26) and (1.27) and solve for the low cost firm's expected first period transfer:

$$\underline{t}_1 = \frac{\gamma}{2}(\underline{\beta} - \underline{c}_1)^2 + \frac{\gamma}{2}(\bar{\beta} - \bar{c}_1)^2 - \frac{\gamma}{2}(\underline{\beta} - \bar{c}_1)^2 + \delta \int_{\mathbb{R}} \underline{U}_2(\rho_2)(\bar{g} - \underline{g})dc_1. \tag{1.28}$$

The first three terms on the right hand side of (1.28) comprise the familiar static transfer; the low cost firm must be compensated for the cost of its effort, and also for the ability to "hide behind" the high cost firm.

In dynamic games, there is an additional component of the low cost firm's first period transfer. Because the density of noise, g, satisfies the monotone likelihood ratio property, the distribution of costs induced by targeting $\bar{c}_1$ first order stochastically dominates the distribution induced by targeting $\underline{c}_1$. Therefore, the low cost firm enjoys a higher expected second period rent when he targets $\bar{c}_1$ than he does when he targets $\underline{c}_1$.[5] The first period transfer must compensate him for this opportunity cost to induce him to target $\underline{c}_1$.

In a deterministic setting, unless the the firm cares little about the future (i.e., the firm heavily discount future payoffs), this additional component of the low cost firm's first period transfer can make it impossible to induce a separating equilibrium. To see this, recall that in a deterministic setting, the firm has perfect control over the project's cost. Suppose the regulator's contract specifies that the high and low cost firms complete the project at different cost levels. If the firm accepts such a contract, his actions perfectly reveal his type to the regulator; information revelation in a deterministic separating equilibrium is an "all-or-nothing" proposition.

Thus, when the the low cost firm follows the equilibrium in the first period, the regulator

[5] That is, because g satisfies the monotone likelihood ratio property, $\int_{\mathbb{R}} \underline{U}_2(\rho_2)(\bar{g} - \underline{g})dc_1 > 0$.

believes with probability one that she is contracting with the low cost type in the second period, and he is held to his reservation utility. Further, when the low cost firm takes out-of-equilibrium actions in the first period and mimics the high cost firm, at the beginning of the second period the regulator believes the firm to be the high cost type. In this case the low cost firm enjoys his highest possible second period information rent, $\underline{U}_2(0)$. To induce him to target $\underline{c}_1$, the principal must increase the low cost firm's first period transfer by $\delta\underline{U}_2(0)$.

This rationale changes in a stochastic setting. First, simply by following the equilibrium and targeting $\underline{c}_1$ in the first period, the low cost firm enjoys expected second period rent

$$\int_{\mathbb{R}} \underline{U}_2(\rho_2)\underline{g}dc_1 > 0. \tag{1.29}$$

Second, the low cost firm's gains from mimicking the high cost firm are diminished. Suppose the low cost firm deviates and targets $\bar{c}_1$ in the second period. The corresponding density of first period costs is $\overline{g}$, so that the low cost firm's expected second period rent from targeting $\bar{c}_1$ is

$$\int_{\mathbb{R}} \underline{U}_2(\rho_2)\bar{g}dc_1 < \int_{\mathbb{R}} \underline{U}_2(0)\bar{g}dc_1 = \underline{U}_2(0). \tag{1.30}$$

Therefore, the additional component of the low cost firm's first period transfer is smaller in a stochastic setting than it is in a deterministic environment.

To proceed with the principal's first period problem, consider the following assumption:

Assumption 1.1. *The single crossing property holds in the first period. That is,*

$$\begin{aligned} \gamma\big(\bar{\beta} - c\big) &\geq \gamma\big(\underline{\beta} - c\big) + \delta \int_{\mathbb{R}} \frac{d\underline{U}_2}{d\rho_2}\frac{d\rho_2}{dc_1} g(c_1 - c)dc_1 \\ \Longrightarrow \gamma\Delta\beta &\geq \delta \int_{\mathbb{R}} \frac{d\underline{U}_2}{d\rho_2}\frac{d\rho_2}{dc_1} g(c_1 - c)dc_1. \end{aligned} \tag{1.31}$$

The single crossing assumption guarantees a regular first period problem by ensuring that the high cost type's marginal cost of decreasing the cost target c is higher than the low cost type's marginal cost of decreasing the cost target for every c. From (1.31), this condition is satisfied when $\frac{d\rho_2}{dc_1}$ is small, i.e. when the posterior beliefs are not too sensitive to changes in first period cost. Since the magnitude of $\frac{d\rho_2}{dc_1}$ depends on the slope of the density, and the slope of the density goes to zero when the variance is large, this condition is satisfied in sufficiently noisy environments. The single crossing condition is also more likely to be satisfied when the difference between the low and high cost firm's intrinsic cost levels, $\Delta\beta$, is large.

Proposition 1.2. *The regulator's full first period problem is given by*

$$\max_{\underline{c}_1, \bar{c}_1} \; S - \rho \left[(1+\lambda) \left(\underline{c}_1 + \frac{\gamma}{2}(\underline{\beta} - \underline{c}_1)^2 \right) + \lambda \left(\frac{\gamma}{2}(\bar{\beta} - \bar{c}_1)^2 - \frac{\gamma}{2}(\underline{\beta} - \bar{c}_1)^2 + \delta \int_{\mathbb{R}} \underline{U}_2(\rho_2)(\bar{g} - \underline{g}) dc_1 \right) \right]$$
$$- (1-\rho)(1+\lambda) \left(\bar{c}_1 + \frac{\gamma}{2}(\bar{\beta} - \bar{c}_1)^2 \right) + \delta E[W_2(\rho_2)], \tag{1.32}$$

where $E[W_2(\rho_2)]$ is given in (1.25). The first order conditions imply the following first period efforts (and cost targets):

$$\underline{e}_1 = \underline{\beta} - \underline{c}_1 = \frac{1}{\gamma} + \frac{\delta}{\gamma\rho(1+\lambda)} \frac{d}{d\underline{c}_1} \left[\rho\lambda \int_{\mathbb{R}} \underline{U}_2(\rho_2)(\bar{g} - \underline{g}) dc_1 - E[W_2] \right], \tag{1.33}$$

and

$$\bar{e}_1 = \bar{\beta} - \bar{c}_1 = \frac{1}{\gamma} - \frac{\rho\lambda}{(1-\rho)(1+\lambda)} \Delta\beta$$
$$+ \frac{\delta}{\gamma(1-\rho)(1+\lambda)} \frac{d}{d\bar{c}_1} \left[\rho\lambda \int_{\mathbb{R}} \underline{U}_2(\rho_2)(\bar{g} - \underline{g}) dc_1 - E[W_2] \right]. \tag{1.34}$$

If the regulator were able to commit to the first and second period cost targets at the outset of her relationship with the firm, she would implement the same contract in each period. In periods one and two, the low cost agent exerts the first best level of effort,

$$\underline{e}^c = \underline{e}^* = \frac{1}{\gamma}. \tag{1.35}$$

The high cost firm's effort distortion remains the same in periods one and two:

$$\bar{e}^c = \frac{1}{\gamma} - \frac{\rho\lambda}{(1-\rho)(1+\lambda)}\Delta\beta. \tag{1.36}$$

Comparing (1.35) to (1.33) and (1.36) to (1.34), one can see that each type of firm's effort is distorted away from the commitment optimum. Whether the low cost firm exerts more or less effort than in the commitment optimum depends on how the additional component of the low cost firm's first period transfer and expected second period welfare change with the low cost firm's first period cost target.

In particular, if

$$\frac{d}{d\underline{c}_1}\left[\rho\lambda\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1 - E[W_2]\right] < 0, \tag{1.37}$$

the low cost firm exerts less effort in the first period than he does in the second period. To see this, recall that the second period game is static. In a static game, the low cost firm exerts the first best effort. The low cost firm also exerts the first best effort in every period when the principal can commit. Therefore, if the low cost firm's first period effort, given in (1.33), is less than the commitment effort given in (1.35), then his first period effort is lower than his effort in the second period.

This case is of particular interest in light of the discussion of the ratchet effect in the introduction. If $\underline{e}_1 < \underline{e}_2$, then the theoretical predictions of this paper match with anecdotal, experimental and empirical evidence which shows that high ability agents decrease their effort at the beginning of their relationship with a principal to maintain information rents in the future.

1.4.1 Signal dampening

Recall that the low cost firm's expected second period rent is higher when he targets $\overline{c}_1$ that it is when he targets $\underline{c}_1$. The additional component of the low cost type's first period transfer,

$$\delta \int_{\mathbb{R}} \underline{U}_2(\rho_2)(\bar{g} - \underline{g}) dc_1, \tag{1.38}$$

compensates him for this difference in expected second period rents. Without this additional component, the principal cannot induce the low cost firm to target $\underline{c}_1$. Clearly, the larger is (1.38), the larger is the low cost firm's first period transfer, given in (1.28). This subsection demonstrates that the principal can decrease (1.38), and thus decrease the low cost firm's expected first period transfer, by reducing the distance between the first period cost targets.

The intuition for this argument is simple. Because the density of noise satisfies the monotone likelihood ratio property, the principal's belief that the firm is the low cost type is monotone decreasing in the first period cost realization. That is, the higher is the first period cost, the lower is the principal's second period belief that the firm is the low cost type.

The lower is the principal's belief that the firm is the low cost type, the more effort the high cost firm exerts in the second period. The more effort that the high cost firm exerts, the higher is the low cost firm's information rent. Thus, the less the principal's second period beliefs change depending on which cost level the firm targets, the lower is the low cost firm's incentive to mimic the high cost firm. To see this, consider Figure 1.1.

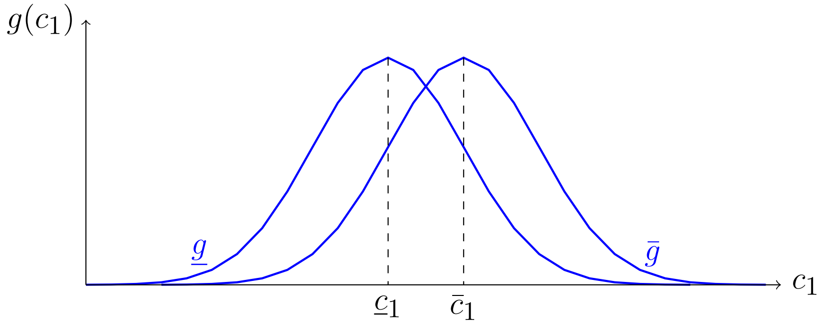

Figure 1.1: The probability density of costs depends on the agent's effort choice

When the firm targets $\underline{c}_1$, the density of first period costs is given by $\underline{g}$ in Figure 1.1. Similarly, when the firm targets $\bar{c}_1$, the density of first period costs is $\overline{g}$. The closer together are $\underline{c}_1$ and $\bar{c}_1$, the closer together are the values of $\underline{g}$ and $\overline{g}$ for any given first period cost realization. The closer together are the values of $\underline{g}$ and $\overline{g}$, the closer second period beliefs, given in (1.5), are to the prior, ρ.

The less the regulator updates her beliefs for any given first period cost realization, the closer is the low cost firm's expected second period rent from targeting $\underline{c}_1$ compared to when he deviates and targets $\bar{c}_1$. This decreases the low cost type's incentives to mimic the high cost type in the first period, which reduces the low cost type's first period transfer. and thus alleviates the first period incentive problem.

The following proposition formalizes this logic by, for the time being, abstracting from the impacts of the first period contract on expected second period welfare. The proof makes use of the connection between effort and cost targets; an increased cost target implies a decrease in effort, and vice-versa. The proof formalizes the intuition that the regulator can decrease the low cost firm's first period transfer by decreasing the distance between $\bar{c}_1$ and

$\underline{c}_1$. To do this, the proof shows that the first period transfer is decreasing in $\underline{c}_1$ and increasing in $\bar{c}_1$. This equilibrium transfer effect decreases (increases) the low cost (high cost) type's equilibrium first period effort.

Proposition 1.3. *The effect of the dynamic portion of the low cost firm's first period transfer is to decrease (increase) the low cost (high cost) firm's first period effort. That is,*

$$\frac{d}{d\underline{c}_1}\left[\rho\lambda\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1\right] < 0, \tag{1.39}$$

and

$$\frac{d}{d\bar{c}_1}\left[\rho\lambda\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1\right] > 0. \tag{1.40}$$

The proof of Proposition 1.3, which is found in Appendix A, establishes that even though the regulator cannot commit to ignore information she learns about the firm when designing the second period contract, in a stochastic environment the regulator can commit to learn less via her choice of first period cost targets. Doing so preserves the low cost firm's equilibrium expected second period rent and decreases his gains from deviation, which in turn decreases his first period transfer, alleviating the dynamic incentive problem.

Tying cost targets to efforts also allows a discussion of how the ratchet effect behaves in a stochastic setting versus a deterministic one. In a deterministic separating equilibrium, the high cost type has his effort decreased over time, while the low cost type always exerts the first best effort.[6] As Proposition 1.3 shows, and as the above intuition argues, in a stochastic setting the regulator distorts the efforts of both types of firm in the first period, as opposed to just the high cost firm. In particular, to decrease the low cost firm's first period transfer, the principal decreases the low cost type's effort, and increases the high cost type's effort,

[6] See Laffont and Tirole (1993).

relative to the commitment optimum.

1.4.2 Experimentation

Proposition 1.3 establishes that the regulator has an incentive to restrict how much information she gathers about the firm. However, an opposing incentive exist as well. The more the regulator learns about the firm's type by observing the first period project cost, the better she can tailor the second period contract to the firm's type. The stronger is the regulator's belief that the firm is the low cost type (i.e., the closer ρ_2 is to one), the lower is the high cost agent's effort. This extracts rent from the low cost firm in the second period. The stronger is the regulator's belief that the firm is the high cost type (i.e., the closer ρ_2 is to zero), the higher is the high cost type's cost-reducing effort.

Thus, the better is the principal's information in the second period, the more accurate is the high cost firm's effort distortion in the second period. This improves expected second period welfare by either inducing more cost-reducing effort from the high cost firm or extracting more rent from the low cost firm. The following lemma establishes that information about the firm's type is valuable to the regulator in the second period.[7]

Lemma 1.1. *Information is valuable. That is, expected second period welfare is convex in second period beliefs:*

$$\frac{d^2 W_2(\rho_2)}{d\rho_2^2} > 0. \tag{1.41}$$

The proof of Lemma 1.1 is a straightforward envelope theorem argument, and is relegated to Appendix A. Given that information is valuable, one can show that the regulator increases expected second period welfare, $E[W_2(\rho_2)]$, by increasing the distance between first period

[7] Information is valuable in the sense of Blackwell (1951).

cost targets.

To see the intuition for this result, return attention to Figure 1.1. As the distance between first period cost targets grows, so does the difference between the value of $\underline{g}$ and $\overline{g}$ for any given first period cost realization. The further apart are the values of $\underline{g}$ and $\overline{g}$, the more the regulator updates her prior beliefs for any given first period cost realization.

Thus, the information asymmetry between the regulator and the firm in the second period diminishes with the distance between first period cost targets. Since welfare distortions in the second period arise because of asymmetric information, an increase in the distance between $\underline{c}_1$ and $\bar{c}_1$ increases expected second period welfare.

This incentive to manipulate first period cost targets to increase how much the principal learns about the agent's type can be interpreted in terms of equilibrium first period efforts. As the following proposition shows, the principal increases expected second period welfare by increasing the low cost firm's effort, and decreasing the high cost firm's effort, relative to the commitment optimum.

Proposition 1.4. *The effect of expected second period welfare is to increase (decrease) the low cost (high cost) firm's first period effort. That is,*

$$\frac{dE[W_2(\rho_2)]}{d\underline{c}_1} < 0, \tag{1.42}$$

and

$$\frac{dE[W_2(\rho_2)]}{d\bar{c}_1} > 0 \tag{1.43}$$

The proof of Proposition 1.4 (found in Appendix A) establishes that the principal increases expected second welfare by increasing the distance between the first period cost targets. Since the game ends after the second period interaction, the only welfare distortions in the second period arise because of the presence of asymmetric information (i.e. there are no dynamic considerations as there are in the first period). Thus, any measures the regulator can take to decrease the information asymmetry in the first period increase expected second period welfare.

1.5 Equilibrium ratchet effect

The analysis has shown that two opposing incentives determine the optimal first period contract. To decrease the low cost firm's first period transfer, the regulator must decrease the distance between the first period cost targets, and restrict how much she learns about the firm's type. To increase expected second period welfare, the regulator must increase the distance between first period cost targets, and increase how much she learns about the firm's type.

To determine the combined effect of these competing incentives on the first period cost targets, consider the following re-formulation of the regulator's first period problem:

$$\begin{aligned}\max_{\underline{c}_1,\bar{c}_1} \quad & S-\rho\left[(1+\lambda)\left(\underline{c}_1+\frac{\gamma}{2}(\underline{\beta}-\underline{c}_1)^2\right)+\lambda\left(\frac{\gamma}{2}(\bar{\beta}-\bar{c}_1)^2-\frac{\gamma}{2}(\underline{\beta}-\bar{c}_1)^2\right)\right] \\ & -(1-\rho)(1+\lambda)\Big(\bar{c}_1+\frac{\gamma}{2}(\bar{\beta}-\bar{c}_1)^2\Big)+\delta\left[\rho\underline{w}^{FB}+(1-\rho)\left(\overline{w}^{FB}-\frac{1+\lambda}{2\gamma}\right)\right] \\ & +\delta\int_{-\infty}^{c_1^0}\left\{(1-\rho)(1+\lambda)\bar{e}_2^0-(1+\lambda-\rho)\frac{\gamma}{2}(\bar{e}_2^0)^2\right\}\bar{g}dc_1 \\ & +\delta\int_{c_1^0}^{\infty}\left\{(1-\rho)(1+\lambda)\bar{e}_2-(1+\lambda-\rho)\frac{\gamma}{2}\bar{e}_2^2\ +\rho\lambda\frac{\gamma}{2}(\bar{e}_2-\Delta\beta)^2\right\}\bar{g}dc_1. \qquad (1.44)\end{aligned}$$

Note that $\underline{w}^{FB}=S-(1+\lambda)\left(\underline{\beta}-\frac{1}{2\gamma}\right)$ and $\overline{w}^{FB}=S-(1+\lambda)\left(\bar{\beta}-\frac{1}{2\gamma}\right)$ are the first best welfare for the low and high cost firm, respectively.

In (1.44), the expected transfers have already been substituted using the low cost firm's incentive constraint and the high cost firm's participation constraint. The second period welfare distortions (how much rent to leave the low cost firm and how much effort to induce in the high cost firm) are captured by the two integrals. Recall that the high cost firm's second period effort determines how much rent is left to the low cost firm. Now, define

$$A := (1-\rho)(1+\lambda)\bar{e}_2^0 - (1+\lambda-\rho)\frac{\gamma}{2}(\bar{e}_2^0)^2, \tag{1.45}$$

and

$$B := (1-\rho)(1+\lambda)\bar{e}_2 - (1+\lambda-\rho)\frac{\gamma}{2}\bar{e}_2^2 + \rho\lambda\frac{\gamma}{2}(\bar{e}_2 - \Delta\beta)^2. \tag{1.46}$$

The first order conditions of this problem imply the following effort levels for the low and high cost firm:

$$\underline{e}_1 = \underline{\beta} - \underline{c}_1 = \frac{1}{\gamma} - \frac{\delta}{\rho(1+\lambda)\gamma}\frac{d}{d\underline{c}_1}\left[\int_{-\infty}^{c_1^0} A\,\bar{g}\,dc_1 + \int_{c_1^0}^{\infty} B\,\bar{g}\,dc_1\right], \tag{1.47}$$

$$\begin{aligned}\bar{e}_1 = \bar{\beta} - \bar{c}_1 =& \frac{1}{\gamma} - \frac{\rho\lambda}{(1-\rho)(1+\lambda)}\Delta\beta \\ & - \frac{\delta}{(1-\rho)(1+\lambda)\gamma}\frac{d}{d\bar{c}_1}\left[\int_{-\infty}^{c_1^0} A\,\bar{g}\,dc_1 + \int_{c_1^0}^{\infty} B\,\bar{g}\,dc_1\right].\end{aligned} \tag{1.48}$$

Again, the equilibrium efforts in (1.47) and (1.48) are distorted relative to the commitment optimum targets in (1.35) and (1.36). The overall effect of the first period contract is to restrict how much the regulator learns about the firm's type if $\bar{e}^c < \bar{e}_1 < \underline{e}_1 < \underline{e}^c$, and to increase learning if $\bar{e}_1 < \bar{e}^c < \underline{e}^c < \underline{e}_1$.

When the distribution of noise is uniform, the overall effect of the first period contract is to decrease the distance between performance targets, relative to the commitment optimum, and restrict learning. This implies that the high-ability agent has its effort increased over the course of his interaction with the principal.[8] However, this result depends on the distribution of noise being uniform. Here, this result is extended to show that when the distribution of noise is general, the net effect of the two competing incentives is to restrict learning; that is, the low cost firm has his effort increased over the course of his interaction with the regulator.

This result that an agent with favorable private information increases his effort over time is appealing because it fits with anecdotal, experimental, and empirical evidence of the ratchet effect. Anecdotal evidence of piece-rate factory workers documented that skilled workers learned to restrict their output in order to avoid either an increase in their output quotas or a decrease in their piece rates.[9] In experimental settings that study two-period principal agent interactions, high ability workers restrict their output (reduce their effort) in the first period to maintain a second period information rent.[10] Empirical studies of the ratchet effect show that teachers reduce their effort on improving student's standardized test scores when their compensation in the future depends on their student's scores today.[11]

With this discussion on the relevance of the ratchet effect in mind, consider the following proposition:

[8]For the uniform noise case, see Jeitschko et al. (2002) and for the general noise case, see Jeitschko and Mirman (2002).

[9]See Matthewson (1931), Clawson (1980), Montgomery (1979) and Roy (1952).

[10]See Charness et al. (2011) and Cardella and Depew (2018).

[11]See Macartney (2016).

Proposition 1.5. ***The Ratchet Effect:*** *If the low cost firm's second period effort from mimicking the high cost firm is positive for all first period cost realizations* $c_1 \geq \underline{c}_1$*, then the low cost firm has his effort increased over the course of the relationship with the regulator. That is,*

$$\frac{d}{d\underline{c}_1}\left[\int_{-\infty}^{c_1^0} A\,\bar{g}\,dc_1 + \int_{c_1^0}^{\infty} B\,\bar{g}\,dc_1\right] > 0, \tag{1.49}$$

$$\frac{d}{d\bar{c}_1}\left[\int_{-\infty}^{c_1^0} A\,\bar{g}\,dc_1 + \int_{c_1^0}^{\infty} B\,\bar{g}\,dc_1\right] < 0. \tag{1.50}$$

The proof of Proposition 1.5 is given in Appendix A. The important implication of Proposition 1.5 is that the low cost firm's first period effort, given in (1.47), is less than his effort when the regulator can commit, (1.35). Since the low cost firm exerts the first best effort in the first period when the principal can commit, and he exerts the first best effort in the second period regardless of the principal's commitment powers, this implies that the low cost firm's effort increases over time.

Since the low cost firm's first period effort is less than in the commitment optimum and the high cost firm's effort is greater than in the commitment optimum, the first period cost targets are closer together than the commitment optimum targets. Therefore, the optimal first period contract favors reducing the first period transfer to the low cost firm at the expense of having worse information about the firm's type in the second period.

Proposition 1.5 requires that the low cost firm's effort from mimicking the high cost firm in the second period be positive for all first period cost realizations greater than the low cost firm's first period cost target. Recall from the discussion of the second period game that there exists a unique first period cost realization, c_1^0, such that for all $c_1 \leq c_1^0$, the low cost firm's effort from mimicking the high cost firm in the second period is negative, and for

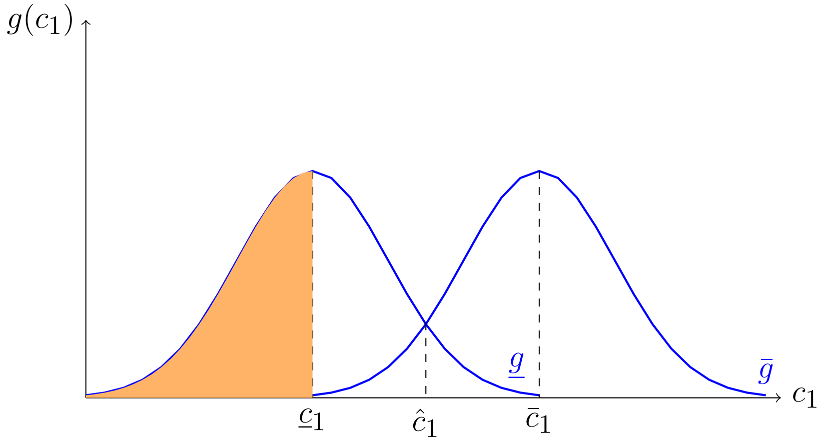

Figure 1.2: If c_1^0 lies in the shaded region, the low cost firm has his effort increased over time.

all $c_1 > c_1^0$ the low cost firm exerts positive effort to mimic the high cost firm. Therefore, Proposition 1.5 requires that c_1^0 be less than or equal to the low cost firm's first period cost target.

Figure 1.2 illustrates the restriction that Proposition 1.5 places on c_1^0, which we consider to be natural. Suppose that $c_1^0 > \underline{c}_1$. This implies that for some cost realizations greater than the low cost firm's cost target, ρ_2 is close enough to one that the high cost firm's second period effort is close to zero. When the high cost firm's effort is close to zero, the low cost firm has to increase costs above its intrinsic cost level, $\underline{\beta}$, to mimic the high cost firm.

Under the conditions outlined in Proposition 1.5, the value of information is decreased in a repeated relationship; not only is the regulator content to have imperfect information in the second period, but she chooses to learn less than she could by implementing the commitment optimum. This is because the benefit of better information in the second period does not outweigh the concomitant increase in the low cost type's expected first period transfer.

1.6 Conclusion

In this two-period model of regulation, the regulator and the firm contract over the completion of a socially valuable project. The firm has private information about its intrinsic cost level, which can be high or low, and has imperfect control over the project's final cost (costs are stochastic). In this setting, the regulator determines how much information she gathers about the firm's type via her choice of first period cost targets.

The regulator can gather more information about the firm by increasing the distance between first period cost targets. The better the regulator's information is about the firm's type in the second period, the higher is expected second period welfare. Conversely, the regulator gathers less information about the firm by decreasing the distance between first period cost targets. The less the regulator learns about the firm's type, the higher is the low cost firm's equilibrium expected second period rent, and the lower is its benefit from mimicking the high cost firm. Thus, by decreasing the distance between first period cost targets, the regulator decreases the low cost firm's first period transfer.

Given a natural restriction on the regulator's second period beliefs, the net effect of the first period contract is to decrease the distance between the first period cost targets. Thus, the regulator's desire to reduce the first period transfer is stronger than her desire to improve expected second period welfare.

This implies that the low cost type exerts less than the first-best effort in the first period, and has his effort ratcheted up over the course of his interaction with the regulator. Anecdotal, experimental and empirical evidence of the ratchet effect suggests that agents with favorable private information preserve their future information rents by taking actions to keep this information private. Thus, the prediction that the low cost firm increases his effort over time aligns closely with observed repeated principal-agent interactions.

Chapter 2

Repeated Short-Term Contracting with Correlated Types and Noisy Observable Outcomes

2.1 Introduction

In many principal-agent interactions, the agent performs the same task for the principal over time. A car salesman sells cars year after year and a science teacher covers the same material, at the same grade level, year after year. In procurement, a firm provides the same good or service to a government agency over time.

In these various settings, the agent's performance on a given task is positively impacted by both his inherent skill and the amount of effort he exerts. No matter how talented and diligent the agent is, however, factors outside his control can impact his performance. Suppose, for example, that the science teacher is judged on his student's standardized test scores. Clearly, he cannot control how much sleep his students get the night before the test or whether some students are sick on the day of the exam. Therefore, the student's test scores are a noisy indicator of the teacher's effort and ability.

Contracts between principal and an agent in which compensation is based (at least in part) on a noisy signal of performance are commonplace. In some states (see [31]), teachers receive a bonus if the average of their student's standardized test scores exceed some pre-specified benchmark. Similarly, a car salesman may receive a bonus if his monthly sales exceed a quota. Often, procurement contracts call for cost sharing between the firm and government agency for cost overruns.

Jeitschko et al. (2002) and Jeitschko and Withers (2018) study two-period principal-agent interactions in which the agent's reward depends solely on a noisy, observable outcome. In both papers, the principal is unable to commit to long term contracts. Therefore, the principal updates her beliefs about the agent's type after observing the noisy outcome, and uses this information when designing the second period contract. The key insight from these papers is that the principal's first period contract choice impacts how much information she gathers about the agent's private characteristic. The principal can design the first period contract to increase how much she learns about the agent's private information, which increases her expected second period payoff, or she can design the first period contract to reduce how much she learns about the agent's private information, which reduces the good agent's first period transfer.

Jeitschko and Withers (2018) shows that the optimal first period contract favors reducing the good agent's first period transfer at the expense of reducing the principal's expected second period payoff, regardless of the distribution of noise.[1] One key feature of these models, however, is that the agent's private information is fixed over time; if the agent has high ability at the beginning of the relationship, he is guaranteed to have high ability at the end of the relationship as well.

[1] It is only assumed that the density satisfies the monotone likelihood ratio property. This extends the results of Jeitschko et al. (2002) and Jeitschko and Mirman (2002).

In real-world settings, however, it may not be realistic to think about the agent's type as fixed. Consumers' preferences change over time, a taxpayer's ability to earn income changes over time, and firms that have industry leading production technology today can be surpassed in the future if a rival makes an innovation.

This paper considers a two period model in which an agent produces output for a principal in each period. The agent's inherent productivity (type) is positively correlated across periods. The principal observes a noisy signal of the agent's performance at the end of the first period. From this signal, the principal learns something about what the agent's type was in the first period, and she uses this information when designing the second period contract. As in Jeitschko et al. (2002), Jeitschko and Mirman (2002) and Jeitschko and Withers (2018), the principal has competing incentives to increase her expected second period payoff and reduce the good agent's first period transfer. Unlike these papers, however, a third incentive is introduced: with some positive probability, the agent with low ability in the first period will have high ability in the second period. Thus, the agent with low ability in the first period earns earns an expected second period rent. Therefore, the principal must consider the impact of the first period contract on the low productivity agent's first period transfer.

The following results hold regardless of the degree of positive correlation. First, the principal reduces the high productivity agent's first period payment by designing the first period contract to restrict how much she learns about the agent's first period type. Second, the principal increases her expected second period payoff, and decreases the low productivity agent's first period transfer, by designing the first period contract to increase how much she learns about the agent's first period type. Third, the principal reduces the total expected first period transfer (the prior-weighted sum of the high and low productivity agent's first period transfers) by designing the first period contract to restrict how much she learns about the agent's private information. Lastly, sufficient conditions are given for the optimal first period contract to favor the reduction of the total expected first period transfer at the expense of having worse information in the second period.

In addition to the literature on principal-agent contracting when the contractible outcome is stochastic, this paper is related to three strands of dynamic principal agent literature. First, this paper is closely related to theoretcial studies of the ratchet effect. The ratchet effect arises in dynamic pricipal agent interactions in which the principal cannot commit to future incentive schemes. The intuition behind the ratchet effect is that the agent can avoid more demanding incentives in the future by reducing his effort in the present. In Weitzman (1980), the ratchet effect arises because the agent's present-day performance target depends explicitly on his past output history. In Freixas, Guesnerie and Tirole (1985), Laffont and Tirole (1987) and Laffont and Tirole (1988), asymmetric information drives the ratchet effect. As in the current paper, the agent's performance on a project is a function of his inherent ability and his performance-enhancing effort. However, these papers assume that the project's outcome is perfectly determined by the agent's choice of effort. For this reason, the dynamic predictions of these papers differ markedly from the predictions of the dynamic-stochastic literature (see Jeitschko and Withers (2018) for further discussion).

Second, this paper is also related to dynamic principal-agent models in which the agent's type is stochastic. One of the first such papers is Baron and Besanko (1984), which studies a multi-period relationship between a regulator and a firm, in which the firm's private cost characteristics may change over time. Laffont and Tirole (1996) examine market based and regulatory solutions for investing in pollution reducing technologies when a firm's "valuation for polluting" changes over time. Battaglini (2005) studies an infinite-horizon pricing problem in which the consumer's preferences evolve according to a Markov process. Battaglini (2007) considers the optimal renegotiation-proof contract in a two period model of procurement in which the agent's type is positively correlated over time. Battaglini and Coate (2008) study optimal income taxation in which an individual's income generating abilities may change over time.

Lastly, this paper is related to a growing dynamic mechanism design literature. Recently, Athey and Segal (2013) and Pavan et al. (2014) study efficient and revenue maximizing dynamic mechanisms, respectively, when the agent's private information is allowed to change over time. An important difference between these papers and the current paper is that the principal has the power to commit to future mechanisms (for a survey on dynamic mechanism design when the principal can commit, see Bergemann and Valimaki (2017)). Skreta (2015) and Gerardi and Maestri (2017) study dynamic mechanisms in which the principal has limited commitment powers, but the agent's private information is fixed.

Most closely related to the current paper is Deb and Said (2015). The authors study a monopolist that faces two cohorts of buyers; consumption occurs only at the end of the second period, and the principal cannot differentiate between second cohort buyers and first cohort buyers who did not agree to a contract in the first period. While the principal can commit to a contract in the first period that specifies the terms of consumption in period two, she cannot commit in the first period to the contract that will be offered in the second period. The setting is dynamic in the sense that the preferences of buyers in the first cohort may change between the first and second periods.[2] The principal finds it optimal to induce some subset of first period buyers to delay their purchase until the second period.

While the principal has limited commitment powers and the agent's private information

[2]Deb and Said (2015) build on the work of Courty and Li (2000).

may change over time, the nature of the principal's problem in Deb and Said (2015) is quite different from the principal's problem in the current paper. In the current paper, the principal interacts with the same agent in each period, and the agent produces the same good for the principal in each period. As in models of the ratchet effect, the principal's concern is to determine how much information she gathers about the agent's private characteristic via her first period contract choice.

2.2 Model

The agent produces output for the principal in time periods one and two. The agent's output in period t is given by

$$y_t = \theta_t e_t + \varepsilon_t, \quad t = 1,\, 2, \tag{2.1}$$

and depends on the agent's ability, θ_t, his effort, e_t, and a zero mean noise term, ε_t. Effort is positive ($e_t \in \mathbb{R}_+$), so the agent improves his expected output in each period by increasing his effort. The agent's productivity can be low or high ($\theta_t \in \{\underline{\theta},\, \overline{\theta}\}$ with $0 < \underline{\theta} < \overline{\theta}$), where type $\overline{\theta}$ is the high productivity or high ability type, and type $\underline{\theta}$ is the low productivity or low ability type. The noise term ε_t is assumed to be distributed uniformly on $[-\eta,\, \eta]$. Prior to her interaction with the agent, the principal believes that the firm has high productivity in the first period with probability $\rho \in (0, 1)$.

The agent's type is positively correlated over time. Thus, with probability $\alpha \in [1/2,\, 1]$ the agent's type in the second period is the same as his type in the first period, and with probability $1 - \alpha$, the agent's type switches between the first and second periods. Therefore, $P(\theta_2 = \overline{\theta} \big|\, \theta_1 = \overline{\theta}) = P(\theta_2 = \underline{\theta} |\, \theta_1 = \underline{\theta}) = \alpha$, and $P(\theta_2 = \underline{\theta} |\, \theta_1 = \overline{\theta}) = P(\theta_2 = \overline{\theta} \big|\, \theta_1 = \underline{\theta}) =$

$1-\alpha$.[3] The special case of $\alpha = 1$ captures the case in which the agent's type is fixed, while $\alpha = 1/2$ capture the cases in which the agent's first and second period types are uncorrelated.

The agent's utility in each period, u_t, is the difference between the transfer he is paid by the principal, $r_t(y_t)$, and his private monetary cost of effort, e_t^2 (i.e., $u_t = r_t(y_t) - e_t^2$). Thus, the agent's expected utility is

$$E[u_t] = E[r_t(y_t) - e_t^2]. \tag{2.2}$$

Notice that the transfer paid to the agent, $r_t(y_t)$, is a function only of observed output. Specifically, this transfer cannot be based on a message from the agent to the principal. This assumption is maintained in order to focus on the impact that imperfect observability has on dynamic incentive problems.[4]

The principal's payoff in each period, $v_t = y_t - r_t$, is simply the difference between the monetary value the principal places on the agent's output and the transfer paid to the agent. Thus, her expected payoff in each period is

$$E[v_t] = E[y_t - r_t(y_t)]. \tag{2.3}$$

The timing of the game is as follows. First, the agent learns his type, θ_1. The principal proposes a payment function, $r_1(y_1)$, that maps from observed output to rewards. If the agent rejects the contract, he receives his outside option of zero. If the agent accepts the contract, he chooses his effort e_1. After his effort has been chosen, ε_1 is realized.[5] The realization of ε_1 determines y_1, which in turn determines the agent's reward and the principal and agent's

[3]This approach to modeling transition probabilities is borrowed from Battaglini (2007).

[4]For a discussion on when it is optimal to base contracts on an additional message from the agent to the principal, see Melumad and Reichelstein (1989).

[5]In each period, effort is chosen before the realization of ε_t; thus, effort is deterministic.

first period payoffs.

At the beginning of the second period, the agent observes his second period type, θ_2. The principal observes the first period output realization, y_1, and updates her beliefs about the agent's first period type using Bayes' rule. She uses her updated beliefs about the agent's first period type, along with the transition probability α, to form her second period beliefs:

$$\rho_2 := P(\theta_2 = \overline{\theta}) = \alpha \cdot P(\theta_1 = \overline{\theta} \big| Y_1 = y_1) + (1-\alpha) \cdot P(\theta_1 = \underline{\theta} \big| Y_1 = y_1). \tag{2.4}$$

Again, the principal offers a reward schedule, $r_2(y_2)$, that maps from second period output realizations to rewards, and the agent accepts or rejects this contract. If the agent accepts the contract, he chooses his effort, and then ε_2 is realized. The principal observes the second period output realization and the principal and agent obtain their second period payoffs. At the end of the second period, the relationship ends. Note that in the following analysis, all proofs are relegated to Appendix B.

2.3 Second period

Suppose the first period equilibrium is such that each type of agent chooses its effort to reach a distinct expected output; that is, suppose that the agent with high productivity chooses his effort in the first period such that $E[y_1] = \overline{y}_1$, while the agent with low productivity in the first period chooses his effort such that $E[y_1] = \underline{y}_1$, and $\overline{y}_1 > \underline{y}_1$. Thus, if the agent has high productivity in the first period, the set of equilibrium output realizations is $y_1 \in [\overline{y}_1 - \eta, \overline{y}_1 + \eta]$, while if the agent has low productivity in the first period, the set of equilibrium output realizations is $y_1 \in [\underline{y}_1 - \eta, \underline{y}_1 + \eta]$.

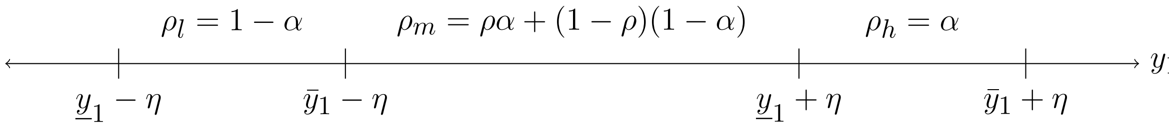

Figure 2.1: Second period beliefs

If the agent targets $\overline{y}_1$ in the first period, the smallest possible first period output realization is $\overline{y}_1 - \eta$. Therefore, when the principal observes output realizations $y_1 < \overline{y}_1 - \eta$, her belief that the firm had low productivity in the first period is equal to one. That is, following an output realization $y_1 < \overline{y}_1 - \eta$, the principal's beliefs about the agent's first period type, updated using Bayes' rule, are as follows: $P(\theta_1 = \underline{\theta} | \, Y_1 = y_1) = 1$ and $P(\theta_1 = \overline{\theta} \big| \, Y_1 = y_1) = 0$. Given these beliefs about the agent's first period type, the principal's second period belief that the agent has high productivity is given by

$$\rho_2 = \alpha \cdot 0 + 1 \cdot (1 - \alpha) = 1 - \alpha =: \rho_l. \tag{2.5}$$

Similarly, if the agent targets $\underline{y}_1$ in the first period, the largest possible output realization is $\underline{y}_1 + \eta$. Therefore, when the principal observes output realizations $y_1 > \underline{y}_1 + \eta$, her belief, updated using Bayes' rule, that the agent had high productivity in the first period is equal to one. Given such beliefs about the agent's first period type, the principal's second period belief that the agent has high productivity is given by

$$\rho_2 = \alpha \cdot 1 + (1 - \alpha) \cdot 0 = \alpha =: \rho_h. \tag{2.6}$$

When the first period output realization is greater than $\bar{y}_1 - \eta$, but less than $\underline{y}_1 + \eta$, the principal is unsure of the agent's first period type. Such output could have been the result of equilibrium behavior by either type of agent in the first period. Because noise is distributed uniformly on $[-\eta, \eta]$, the value of the probability density function is equal to $\frac{1}{2\eta}$ for every potential output realization. Therefore, when the principal observes first period output realizations $y_1 \in [\bar{y}_1 - \eta, \underline{y}_1 + \eta]$, she updates her beliefs about the agent's first period type using Bayes' rule as follows:

$$P(\theta_1 = \overline{\theta} \big| Y_1 = y_1) = \frac{\rho \cdot \frac{1}{2\eta}}{\rho \cdot \frac{1}{2\eta} + (1-\rho) \cdot \frac{1}{2\eta}} = \rho. \tag{2.7}$$

Similarly, $P(\theta_1 = \underline{\theta} | Y_1 = y_1) = 1 - \rho$. Thus, for intermediate output realizations, the principal's second period belief that the firm is the high productivity type is given by

$$\rho_2 = \rho\alpha + (1-\rho)(1-\alpha) =: \rho_m. \tag{2.8}$$

In summary, the principal's second period beliefs that the firm is the high productivity type are given by:

$$\rho_2 = \begin{cases} 1 - \alpha =: \rho_l, & \text{if} \quad y_1 < \bar{y}_1 - \eta, \\ \rho\alpha + (1-\rho)(1-\alpha) =: \rho_m, & \text{if} \quad y_1 \in [\bar{y}_1 - \eta, \underline{y}_1 + \eta], \\ \alpha =: \rho_h, & \text{if} \quad y_1 > \underline{y}_1 + \eta. \end{cases} \tag{2.9}$$

Consider the second period game in which the principal has generic belief ρ_2 that the agent is the high productivity type in the second period. Given these beliefs, the principal's problem is to design a reward schedule that maximizes her expected second period payoff, subject to a set of participation and incentive constraints. This reward schedule, $r_2(y_2)$, is based solely on observed second period output.

Rather than focus on deriving this reward schedule, however, the first and second period analysis focuses on the principal's choice of profit maximizing output targets. In period $t = 1, 2$, the high productivity agent's output target is given by $\overline{y}_t$, and the low productivity agent's output target by $\underline{y}_t$. As long as there exists a reward schedule that satisfies the three following properties in each period, re-framing the principal's problem in such a manner is without loss of generality.

First, the reward schedule must make the high productivity type's expected utility from targeting $\overline{y}_t$ equal to his expected utility from targeting $\underline{y}_t$. Second, it must make the low productivity agent's expected utility from targeting $\underline{y}_t$ equal to his outside option of zero. Lastly, the reward schedule must be such that the high productivity agent's expected utility from targeting $y_t \notin \{\overline{y}_t, \underline{y}_t\}$ is strictly lower than his expected utility from targeting $\overline{y}_t$ or $\underline{y}_t$, and the low productivity firm's expected utility from targeting $\underline{y}_t$ is greater than his expected utility from targeting any other period-t output level.

As Jeitschko et al. (2002) show, one can explicitly derive such a reward schedule when noise follows a uniform distribution. The reward schedule resembles a base pay/bonus incentive scheme. The low productivity agent receives a transfer equal to his cost of effort for all output realizations in $[\underline{y}_t - \eta, \underline{y}_t + \eta]$; this ensures that his participation constraint is satisfied in expectation. The high productivity agent receives a bonus if his output exceeds some cutoff; the cutoff is chosen to satisfy, in expectation, the high productivity types incentive constraint. Since noise is uniform (i.e., has bounded support), the principal can prevent the low productivity agent from shirking by severely punishing output realizations less than $\underline{y}_t - \eta$. The high productivity agent does not wish to reduce his effort, as his expected transfer decreases faster than his disutility of effort. It is shown in the Appendix that such a reward schedule exists when the agent's type is positively correlated.

Given the existence of a reward schedule that implements output targets $\overline{y}_t$ and $\underline{y}_t$ in each period, consider the high productivity agent's second period incentive constraint. The agent's expected output is simply the product of his type and his effort. Thus, the high productivity agent's expected second period output is $E[y_2] = \overline{\theta} \cdot e_2$.[6] Therefore, to target output level $\overline{y}_2$, the high productivity agent chooses his effort to equal $\overline{y}_2/\overline{\theta}$. When the high productivity agent chooses his effort in this manner, the expected transfer is

$$\int_{\overline{y}_2-\eta}^{\overline{y}_2+\eta} r_2(y_2) \cdot \frac{1}{2\eta} dy_2 =: \overline{r}_2, \tag{2.10}$$

where $r_2(y_2)$ is the second period reward schedule. Thus, the high productivity agent's expected utility is simply $\overline{r}_2 - (\overline{y}_2/\overline{\theta})^2$.

When the high productivity agent chooses his effort so that $e_2 = \underline{y}_2/\overline{\theta}$, he targets $\underline{y}_2$. When he targets $\underline{y}_2$, his expected transfer is

$$\int_{\underline{y}_2-\eta}^{\underline{y}_2+\eta} r_2(y_2) \cdot \frac{1}{2\eta} dy_2 =: \underline{r}_2, \tag{2.11}$$

and his expected utility is $\underline{r}_2 - (\underline{y}_2/\overline{\theta})^2$. Through a similar line of reasoning, the low productivity firm's expected utility from targeting $\underline{y}_2$ is $\underline{r}_2 - (\underline{y}_2/\underline{\theta})^2$.

Incentive compatibility for the high productivity type requires that his expected utility from targeting $\overline{y}_2$ be equal to his expected utility from targeting $\underline{y}_2$. Individual rationality for the low productivity type requires that his expected utility from targeting $\underline{y}_2$ be equal to his outside option of zero. The principal's second period problem is to maximize her expected payoff, subject to the high productivity agent's incentive constraint and the low

[6] From (2.1), $E[y_2] = E[\theta_2 \cdot e_2 + \varepsilon_2] = \theta_2 \cdot e_2$.

productivity agent's participation constraint:

$$\begin{aligned}
&\max_{\underline{y}_2, \overline{y}_2} && \rho_2\Big[\overline{y}_2 - \overline{r}_2\Big] + (1-\rho_2)\Big[\underline{y}_2 - \underline{r}_2\Big] && \\
&s.t. && \overline{r}_2 - \left(\frac{\overline{y}_2}{\overline{\theta}}\right)^2 = \underline{r}_2 - \left(\frac{\underline{y}_2}{\overline{\theta}}\right)^2 && (\overline{IC}_2) \\
& && \underline{r}_2 - \left(\frac{\underline{y}_2}{\underline{\theta}}\right)^2 = 0. && (\underline{IR}_2) \qquad (2.12)
\end{aligned}$$

After using the incentive and participation constraints to substitute out for the expected transfers, one derives the second period equilibrium output targets for the high and low productivity agent, respectively:

$$\overline{y}_2 = \frac{\overline{\theta}^2}{2}, \tag{2.13}$$

and

$$\underline{y}_2 = C(\rho_2)\frac{\underline{\theta}^2}{2}, \tag{2.14}$$

where $C(\rho_2) = \frac{1-\rho_2}{1-\rho_2\Theta^2}$, and $\Theta = \underline{\theta}/\overline{\theta}$. The high productivity agent's expected information rent is given by

$$\overline{u}_2(\rho_2) = C^2(\rho_2)\frac{\underline{\theta}^2}{4}\left[1-\Theta^2\right]. \tag{2.15}$$

Observing (2.14) and (2.15), one can see that this second period game exhibits the classic rent-extraction/efficiency trade-off that characterizes static asymmetric information games. As the principal's belief that the agent has high productivity approaches one, the low productivity type has his output target, and thus his effort, reduced to zero ($C(\rho_2)$ goes to zero as ρ_2 goes to one). By reducing the low productivity agent's effort as second period beliefs go to one, the principal extracts rent from the high productivity worker:

$$\frac{d\overline{u}_2(\rho_2)}{d\rho_2} = -\left(\frac{\underline{\theta}\left[1-\Theta^2\right]}{2\left[1-\rho_2\Theta^2\right]}\right)^2 < 0. \tag{2.16}$$

This trade-off has important implications for the first period problem. To see these implications, consider the following lemma:

Lemma 2.1. *The principal's belief that the agent is the high productivity type in the second period is increasing in first period output. That is,* $\rho_l \leq \rho_m \leq \rho_h$.

The proof of Lemma 2.1 is straightforward, given that $\alpha \geq 1/2$. The important implication of Lemma 2.1 is summarized in Corollary 2.1 below:

Corollary 2.1. *The high productivity agent's second period rent is decreasing in first period output. That is,* $\overline{u}_2(\rho_l) > \overline{u}_2(\rho_m) > \overline{u}_2(\rho_h)$.

The proof of Corollary 2.1 follows directly from Lemma 2.1, and the fact that $\frac{d\overline{u}_2(\rho_2)}{d\rho_2} < 0$. To see the importance of Corollary 2.1 in the first period, consider the high productivity agent's first period effort choice. By targeting $\overline{y}_1$ in the first period, the high productivity type's set of possible first period output realizations is $[\overline{y}_1 - \eta, \overline{y}_1 + \eta]$. With probability $\frac{\overline{y}_1 - \underline{y}_1}{2\eta}$, he has a favorable output shock, and the principal learns that the he has high ability in the first period. Her second belief that the worker has high productivity is given by ρ_h. With probability $1 - \frac{\overline{y}_1 - \underline{y}_1}{2\eta}$, the first period high productivity type has an unfavorable output realization. In this case, the principal's belief that the agent has high productivity in the second period is given by ρ_m. The worker only obtains a second period information rent if he remains the high productivity type in the second period. Therefore, the first-period high productivity type's expected second period rent from targeting $\overline{y}_1$ is given by

$$\overline{E}_1[\overline{u}_2(\rho_2)|\,\overline{y}_1] := \alpha\left[\left(1-\frac{\overline{y}_1-\underline{y}_1}{2\eta}\right)\cdot\overline{u}_2(\rho_m)+\left(\frac{\overline{y}_1-\underline{y}_1}{2\eta}\right)\cdot\overline{u}_2(\rho_h)\right]. \tag{2.17}$$

Suppose instead the high productivity agent targets $\underline{y}_1$ in the first period. If he experiences a negative output shock, the principal believes that he was the low productivity agent in the first period. Conditional on remaining the high productivity type in the second period, he enjoys expected second period rent $\overline{u}_2(\rho_l)$. If he experiences a favorable output shock when targeting $\underline{y}_1$, his expected second period rent is $\overline{u}_2(\rho_m)$. From the perspective of the first period, his expected second period rent from targeting $\underline{y}_1$ is

$$\overline{E}_1[\overline{u}_2(\rho_2)|\,\underline{y}_1] := \alpha\left[\left(1-\frac{\overline{y}_1-\underline{y}_1}{2\eta}\right)\cdot\overline{u}_2(\rho_m)+\left(\frac{\overline{y}_1-\underline{y}_1}{2\eta}\right)\cdot\overline{u}_2(\rho_l)\right]. \tag{2.18}$$

Using Corollary 2.1, one can easily verify that the high productivity agent's expected second period rent from targeting $\underline{y}_1$ is larger than his expected second period rent when targeting $\overline{y}_1$. Thus, the high productivity agent benefits from mimicking the low productivity type in the first period. As in the standard ratchet effect literature, the principal must increase his first period transfer by $E[\overline{u}_2(\rho_2)|\,\underline{y}_1] - E[\overline{u}_2(\rho_2)|\,\overline{y}_1]$ to induce him to target $\overline{y}_1$ in the first period.

Unlike the standard ratchet effect literature in which the agent's type is fixed, the agent with low productivity in the first period has an expected second period rent, since with probability $1-\alpha$ he has high ability in the second period. Suppose the agent who has low productivity in the first period chooses to target $\underline{y}_1$ in the first period. With probability $1-\frac{\overline{y}_1-\underline{y}_1}{2\eta}$, he has a favorable output shock, and the principal has beliefs ρ_m that the agent is the high productivity type in the second period. With the opposite probability, he has an unfavorable output shock, and the principal learns that the agent has low productivity in the first period. Therefore, the low productivity agent's expected second period rent from targeting $\underline{y}_1$ in the first period is

$$\underline{E}_1[\overline{u}_2(\rho_2)|\,\underline{y}_1] := (1-\alpha)\left[\left(1 - \frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\cdot \overline{u}_2(\rho_m) + \left(\frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\cdot \overline{u}_2(\rho_l)\right]. \tag{2.19}$$

When the low productivity agent targets $\overline{y}_1$ in the first period, his expected second period rent is

$$\underline{E}_1[\overline{u}_2(\rho_2)|\,\overline{y}_1] := (1-\alpha)\left[\left(1 - \frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\cdot \overline{u}_2(\rho_m) + \left(\frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\cdot \overline{u}_2(\rho_h)\right]. \tag{2.20}$$

It is clear to see that the low productivity agent prefers a first period equilibrium that increases the probability that the principal learns the agent's first period type. The intuition is straightforward; the low productivity agent induces a more favorable distribution of second period output by targeting $\underline{y}_1$ than if he were to mimic the high productivity agent in the first period. The further apart are $\overline{y}_1$ and $\underline{y}_1$, the bigger is the difference between $\underline{E}_1[\overline{u}_2(\rho_2)|\,\underline{y}_1]$ and $\underline{E}_1[\overline{u}_2(\rho_2)|\,\overline{y}_1]$. Therefore, the principal can reduce the low productivity agent's first period transfer by increasing the distance between first period output targets.

Lastly, consider the impacts of the first period contract on the principal's second period payoff, which is given by

$$v_2(\rho_2) = \rho_2\left[\frac{\overline{\theta}^2}{4} - C^2(\rho_2)\frac{\underline{\theta}^2}{4}\left[1-\Theta^2\right]\right] + (1-\rho_2)\left[C(\rho_2)\frac{\underline{\theta}^2}{2}\left[1 - \frac{C(\rho_2)}{2}\right]\right]. \tag{2.21}$$

By increasing the distance between output targets, the principal increases the probability that she learns the agent's first period type. This reduces her uncertainty regarding the agent's second period type, which allows her to reduce the good agent's expected second period rent or reduce the bad agent's effort distortions. Therefore, the principal increases her expected second period payoff by increasing the distance between first period output targets.

2.4 First period

As in the second period, the principal's first period problem is to choose output targets $\overline{y}_1$ and $\underline{y}_1$ to maximize the sum of her first and (discounted) second period expected payoffs,

$$\rho \int_{\overline{y}_1-\eta}^{\overline{y}_1+\eta} [y_1 - r_1(y_1)] \frac{1}{2\eta} dy_1 + (1-\rho) \int_{\underline{y}_1-\eta}^{\underline{y}_1+\eta} [y_1 - r_1(y_1)] \frac{1}{2\eta} dy_1 + \delta E[v_2(\rho_2)], \tag{2.22}$$

where

$$E[v_2(\rho_2)] = \left(\frac{\overline{y}_1 - \underline{y}_1}{2\eta} \right) [\rho v_2(\rho_h) + (1-\rho) v_2(\rho_l) - v_2(\rho_m)] + v_2(\rho_m), \tag{2.23}$$

subject to incentive compatibility and participation constraints, which are derived below.

A well-known problem that arises in dynamic games of asymmetric information is that the low productivity agent's incentive constraint may bind in the first period. As discussed in the previous section, the principal must increase the high productivity agent's first period transfer to induce him to target $\overline{y}_1$ rather than $\underline{y}_1$. When output is deterministic and the agent's type is fixed, the low productivity type is tempted to mimic the high productivity agent in the first period. By doing so, he receives the high productivity agent's large first period transfer, and can walk away from the relationship in the second period.

In the current setting, two factors alleviate this dynamic incentive problem. First, output is noisy, which implies that the principal's learning process is slowed. For intermediate output realizations, the principal is unsure of the agent's first period type. Additionally, the low cost agent receives an expected second period rent, and this rent is higher when targeting $\underline{y}_1$ than if he were to target $\overline{y}_1$. Thus, the dynamic incentive problem is alleviated relative to the deterministic case, and even the stochastic case in which the agent's type is fixed over time. The first period problem proceeds by assuming the low productivity firm's first period incentive constraint is slack. Once the equilibrium output targets are derived, it is verified in Appendix B that for η large enough, this assumption holds.

First, consider the high productivity type's incentive compatibility constraint. When he chooses his effort so that $e_1 = \overline{y}_1/\overline{\theta}$, his expected first period reward is

$$\overline{r}_1 := \int_{\overline{y}_1-\eta}^{\overline{y}_1+\eta} r_1(y_1)\frac{1}{2\eta}dy_1. \tag{2.24}$$

Therefore, his expected first period utility is $\overline{r}_1 - (\overline{y}_1/\overline{\theta})^2$. His expected second period rent from targeting $\overline{y}_1$ is given by (2.17). Suppose instead he chooses his effort so that $e_1 = \underline{y}_1/\overline{\theta}$. In this case, his expected first period reward is

$$\underline{r}_1 := \int_{\underline{y}_1-\eta}^{\underline{y}_1+\eta} r_1(y_1)\frac{1}{2\eta}dy_1, \tag{2.25}$$

and his first period utility is $\underline{r}_1 - (\underline{y}_1/\overline{\theta})^2$. His expected second period rent from targeting $\underline{y}_1$ is given in (2.18). Therefore, incentive compatibility for the high productivity type requires that

$$\overline{r}_1 - \left(\frac{\overline{y}_1}{\overline{\theta}}\right)^2 + \delta E[\overline{u}_2(\rho_2)|\,\overline{y}_1] = \underline{r}_1 - \left(\frac{\underline{y}_1}{\overline{\theta}}\right)^2 + \delta E[\overline{u}_2(\rho_2)|\,\underline{y}_1]. \tag{2.26}$$

From this incentive compatibility constraint, we can derive the high productivity firm's expected first period transfer:

$$\overline{r}_1 = \left(\frac{\overline{y}_1}{\overline{\theta}}\right)^2 + \underline{r}_1 - \left(\frac{\underline{y}_1}{\overline{\theta}}\right)^2 + \delta\alpha\left(\frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\left(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_h)\right). \tag{2.27}$$

Next, consider the high cost firm's participation constraint. By choosing his effort so that $e_1 = \underline{y}_1/\underline{\theta}$, the low cost firm's expected first period utility is $\underline{r}_1 - (\underline{y}_1/\underline{\theta})^2$. His expected second period rent from targeting $\underline{y}_1$ is given in (2.19). Therefore, the low productivity firm's participation constraint is given by

$$\underline{r}_1 - \left(\frac{\underline{y}_1}{\underline{\theta}}\right)^2 + \delta(1-\alpha)\left[\left(\frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\overline{u}_2(1-\alpha) + \left(1 - \frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\overline{u}_2(\rho_2^2)\right] = 0. \tag{2.28}$$

The principal's first period problem is as follows:

$$\begin{aligned}
\max_{\underline{y}_1, \overline{y}_1} \quad & \rho\left[\overline{y}_1 - \overline{r}_1\right] + (1-\rho)\left[\underline{y}_1 - \underline{r}_1\right] + \delta E[v_2(\rho_2)] \\
s.t. \quad & \overline{r}_1 = \left(\frac{\overline{y}_1}{\overline{\theta}}\right)^2 + \underline{r}_1 - \left(\frac{\underline{y}_1}{\overline{\theta}}\right)^2 + \delta\alpha\left(\frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\left(\overline{u}_2(1-\alpha) - \overline{u}_2(\alpha)\right) \\
& \underline{r}_1 = \left(\frac{\underline{y}_1}{\underline{\theta}}\right)^2 - \delta(1-\alpha)\left[\left(\frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\overline{u}_2(1-\alpha) + \left(1 - \frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\overline{u}_2(\rho_2^2)\right].
\end{aligned} \tag{2.29}$$

After using the incentive and participation constraints to eliminate the expected first period transfer from the principal's problem, one obtains the following equilibrium first period output targets for the high and low productivity agent, respectively:

$$\overline{y}_1 = \frac{\overline{\theta}^2}{2} + \frac{1}{\rho}\frac{\overline{\theta}^2}{2}\frac{\delta}{2\eta} \cdot A, \tag{2.30}$$

and

$$\underline{y}_1 = C(\rho)\frac{\underline{\theta}^2}{2} - \frac{1}{1-\rho\Theta^2}\frac{\underline{\theta}^2}{2}\frac{\delta}{2\eta}\cdot A, \tag{2.31}$$

where

$$\begin{aligned} A := & \rho v_2(\rho_h) + (1-\rho)v_2(\rho_l) - v_2(\rho_m) \\ & - (\rho\alpha(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_h)) - (1-\alpha)(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_m)))\,. \end{aligned} \tag{2.32}$$

One of the main results from the stochastic contracting literature in which the agent's type is fixed is that it is optimal for the principal to learn less about the agent's type than she could by setting the first period performance targets equal to the commitment optimum targets.[7] Suppose the principal is able to commit to the second period incentive scheme at the beginning of the interaction with the agent. The first period commitment output targets for the high and low productivity agent, respectively, are given by

$$\overline{y}_1^c := \frac{\overline{\theta}^2}{2} \tag{2.33}$$

and

$$\underline{y}_1^c := C(\rho)\frac{\underline{\theta}^2}{2}. \tag{2.34}$$

Therefore, if $A < 0$, it is optimal for the the principal to learn less about the agent's first period type than she could by choosing the commitment output targets.[8] Before exploring the relationship between the equilibrium first period output targets and the commitment output targets, we examine separately the three effects that determine their relationship:

[7] See Jeitschko et al. (2002) and Jeitschko and Withers (2018).

[8] If $A < 0$, then $\underline{y}_1^c \leq \underline{y}_1 \leq \overline{y}_1 \leq \overline{y}_1^c$.

the principal's desire to reduce the high productivity firm's first period transfer, her desire to reduce the low productivity firm's first period transfer, and her desire to increase her expected second period payoff.

2.4.1 Signal dampening

The principal's choice of first period output targets determines the probability with which she learns the agent's first period type. When the output targets are close together, this probability is small. When the probability that the principal learns the agent's first period type is small, the high productivity agent has little incentive to mimic the low productivity agent in the first period. Therefore, the principal reduces the high productivity agent's first period transfer by setting the first period output targets close together.

To see this, note that when the agent targets $\overline{y}_1$ in the first period, the principal' second period belief that the agent has high productivity is either ρ_m or ρ_h. The closer together are the first period cost targets, the more likely it is that the principal's second period beliefs are given by ρ_m. Since $\overline{u}_2(\rho_m) > \overline{u}_2(\rho_h)$, bringing the first period output targets closer together increases $\overline{E}_1[\overline{u}_2(\rho_2)| \, \overline{y}_1]$.

Similarly, if the high productivity agent targets $\underline{y}_1$ in the first period, the principal's second period belief that the agent has high productivity is either ρ_m or ρ_l. The closer together are the first period output targets, the more likely it is that the principal's second period beliefs are given by ρ_m. Since $\overline{u}_2(\rho_m) < \overline{u}_2(\rho_l)$, bringing the first period output targets closer together decreases $\overline{E}_1[\overline{u}_2(\rho_2)| \, \underline{y}_1]$. Thus, by bringing the first period output targets closer together, the principal decreases the difference between $\overline{E}_1[\overline{u}_2(\rho_2)| \, \underline{y}_1]$ and $\overline{E}_1[\overline{u}_2(\rho_2)| \, \overline{y}_1]$. By reducing the difference in these expected utilities, the principal reduces the high productivity agent's first period transfer.

Proposition 1 below formalizes this argument. To make the statement of Proposition 1 more clear, solve for $\underline{r}_1$ in (2.28) and substitute it into (2.27). The resulting expression for the high productivity agent's first period expected transfer can be decomposed as follows:

$$\overline{r}_1 = \overline{r}_1^S + \overline{r}_1^D, \tag{2.35}$$

where

$$\overline{r}_1^S := \left(\frac{\overline{y}_1}{\overline{\theta}}\right)^2 + \left(\frac{\underline{y}_1}{\underline{\theta}}\right)^2 - \left(\frac{\underline{y}_1}{\overline{\theta}}\right)^2, \tag{2.36}$$

and

$$\begin{aligned}\overline{r}_1^D := \delta\left(\frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\big[\alpha(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_h)) - (1-\alpha)(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_m))\big] \\ - \delta(1-\alpha)\overline{u}_2(\rho_m).\end{aligned} \tag{2.37}$$

The "dynamic" portion of this expected transfer, $\overline{r}_1^D$, represents the difference in expected second period rent that the high productivity agent receives from targeting $\underline{y}_1$ versus targeting $\overline{y}_1$, less the amount that the principal is able to extract from the low productivity agent. If $\overline{r}_1^D$ is increasing in the distance between first period output targets, then the principal can decrease the high productivity firm's expected first period transfer by bringing the first period output targets closer together.

Proposition 2.1. *The principal can decrease the high productivity firm's first period transfer by decreasing the distance between $\overline{y}_1$ and $\underline{y}_1$. That is,*

$$\frac{d\,\overline{r}_1^D}{d(\overline{y}_1 - \underline{y}_1)} > 0. \tag{2.38}$$

It is important to note that Proposition 2.1 is true for any degree of positive correlation between the agent's first and second period types. As discussed above, the benefit of reducing the distance between first period output targets is to decrease the high productivity agent's incentive to mimic the low ability agent in the first period. The cost of reducing the distance between first period output targets is that the principal is able to extract less rent from the low ability agent in the first period.[9]

For the purposes of reducing the high productivity agent's expected first period transfer, the benefit of reducing the distance between output targets outweighs the costs for all levels of positive correlation for two reasons. First, since the low ability agent's incentive constraint is slack, it matters less to the principal to extract rent from the low ability type than it does from the high ability type. Second, since types are positively correlated, an agent with high ability in the first period is more likely to receive a second period rent than an agent with low ability in the first period.

2.4.2 Experimentation

Just as the principal can decrease the probability of learning the agent's first period type by bringing the first period output targets closer together, the opposite is true as well. By spreading $\overline{y}_1$ and $\underline{y}_1$ further apart, the principal increases the probability that she learns the agent's first period productivity parameter.

Because the agent's type is correlated over time, the principal does not have complete information about the his second period type even when she learns the his first period type. Nevertheless, the principal benefits in two ways from learning the agent's first period type. First, the principal can either extract more rent from the high productivity type or induce

[9]This will be discussed in more detail in the following section.

the low productivity type to exert more effort. Second, she is able to reduce the first period expected transfer to the low productivity agent.

First, consider the impact of the distance between first period output targets on the low productivity agent's first period expected transfer. Recall that the low productivity agent chooses his effort to target $\underline{y}_1$ (his incentive constraint is slack). The low productivity agent prefers a lower $\underline{y}_1$ for two reasons. First, a lower output target requires less effort to achieve. Second, the further $\underline{y}_1$ is from $\overline{y}_1$, the larger is the low productivity agent's expected second period rent.

To see this, recall that the low productivity agent receives no second period rent if he remains the low productivity type. Unlike the high productivity agent, therefore, he has no incentive to conceal his first period private information. When he targets $\underline{y}_1$, the principal's second period belief that the agent has high productivity is either ρ_l or ρ_m, depending on whether the first period output realization is less than or greater than $\overline{y}_1 - \eta$. Conditional on becoming the high productivity type in the second period, the low productivity agent prefers a first period equilibrium that places more weight on $\overline{u}_2(\rho_l)$ as opposed to $\overline{u}_2(\rho_m)$. That is, he benefits from a first period equilibrium that increases the probability that the principal learns his first period type.

Proposition 2.2 formalizes this logic. To simplify the statement of the proposition, let

$$\underline{r}_1 = \underline{r}_1^S + \underline{r}_1^D, \tag{2.39}$$

where

$$\underline{r}_1^S := \left(\frac{\underline{y}_1}{\underline{\theta}}\right)^2, \tag{2.40}$$

and

$$\underline{r}_1^D := -\delta(1-\alpha)\left(\frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_m)) - \delta(1-\alpha)\overline{u}_2(\rho_m). \tag{2.41}$$

Proposition 2.2. *The principal can decrease the low productivity firm's first period transfer by moving the first period output targets further apart. That is,*

$$\frac{d\,\underline{r}_1^D}{d(\overline{y}_1 - \underline{y}_1)} < 0. \tag{2.42}$$

The intuition of the result is clear; the portion of the low productivity agent's first period transfer that depends on the distance between first period output targets is decreasing in the distance between those targets. Therefore, the principal decreases the expected first period transfer to the low productivity agent by increasing the distance between $\underline{y}_1$ and $\overline{y}_1$. The economic intuition for this result is discussed in the argument preceding the statement of Proposition 2.2.

Next, consider the effect of the distance between first period output targets on the principal's expected second period payoff. Lemma 2.2 establishes that information about the agent's first period type is valuable to the principal; the more the principal knows about the agent's first period type, the better she can balance the rent extraction/efficiency tradeoff in the second period. Given that information is valuable, Proposition 2.3 establishes that the principal acquires better information about the agent's first period type, and thus increases her expected second period payoff, by increasing the distance between first period output targets.

Lemma 2.2. *Information is valuable to the principal. That is, the principal's second period* *expected payoff is convex in second period beliefs (see Blackwell (1951)):*

$$\frac{d^2 v_2(\rho_2)}{d\rho_2^2} > 0. \tag{2.43}$$

Given Lemma 2.2, it is easy to show that the principal increases her expected second period payoff by increasing the distance between first period output targets.

Proposition 2.3. *The principal increases her expected second period payoff, $E[v_2(\rho_2)]$, by increasing the distance between first period output targets. That is,*

$$\frac{dE[v_2(\rho_2)]}{d(\overline{y}_1 - \underline{y}_1)} > 0. \tag{2.44}$$

Proposition 2.2 and Proposition 2.3 establish that two incentives drive the principal to learn the agent's first period private information. Like the stochastic contracting literature in which the agent's type is fixed, the principal increases her expected second period payoff by learning more about the agent's first period private information. Unlike the aforementioned literature, however, the principal has an incentive to learn more to decrease the first period transfer to the low productivity agent. These two incentives combine with the incentive to decrease the high productivity agent's first period transfer to determine the first period output targets.

2.4.3 Total first period transfer: signal dampening or experimentation?

Both the high and low productivity firm's first period transfers depend on the distance between first period output targets. Proposition 2.1 demonstrates that the principal decreases the high productivity firm's transfer by learning less about the agent's first period type, while Proposition 2.2 demonstrates that the principal decreases the low productivity agent's transfer by learning more about the agent's first period type.

In the principal's first period problem, the high and low productivity firm's first period transfers are weighted by the principal's prior beliefs about the agent's type. Define the "total expected first period transfer," $E[r_1]$, as follows:

$$E[r_1] := \rho \overline{r}_1 + (1-\rho)\underline{r}_1, \tag{2.45}$$

where $\overline{r}_1$ and $\underline{r}_1$ are given in (2.35) and (2.39), respectively. One component of the total expected transfer, $\overline{r}_1$, is increasing in the distance between output targets, while the other component, $\underline{r}_1$, is decreasing in the distance between output targets. This raises the question of how $E[r_1]$ depends on the distance between first period output targets.

To see why this question is interesting, consider the relationship between the optimal first period output targets, given in (2.30) and (2.31), and the commitment output targets, given in (2.33) and (2.34). This relationship depends on the principal's incentives to decrease the high and low productivity agent's first period transfers and her incentive to increase her expected second period payoff. Proposition 2.3 establishes that to increase her expected second period payoff, the principal increases the distance between first period output targets. This result does not depend on the value of the correlation parameter, the level of the prior, or the ratio of the intrinsic productivity levels.

This implies that if, for some values of the primitives, the principal decreases $E[r_1]$ by increasing the distance between output targets, then the optimal first period output targets lie outside the commitment optimum output targets. This result would stand in contrast to the stochastic contracting literature in which the agent's type is fixed; when the agent's type is fixed, a robust result is that the optimal performance targets reveal less information about the agent's type to the principal than she would gather by setting the first period output targets equal to the commitment level. The following proposition, however, shows that this is never the case:

Proposition 2.4. *The principal can decrease the total expected first period transfer by decreasing the distance between the first period output targets. That is,*

$$\frac{dE[r_1]}{d(\overline{y}_1 - \underline{y}_1)} > 0. \tag{2.46}$$

An interesting takeaway from Proposition 2.4 is that, regardless of her prior beliefs, ρ, the principal decreases the total expected transfer by designing the first period output targets to decrease the the first period payment to the high productivity agent. Put another way, if the principal's objective is to reduce the total expected first period transfer, it is never in her interest to design the first period output targets to increase the probability that she learns the agent's first period type, no matter how unlikely the high productivity agent is ex-ante.

2.5 Equilibrium rent preservation

With the exception of subsection 2.4.3, the three incentives that determine the distance between the first period output targets have been considered in isolation. First, the principal decreases the high productivity agent's first period transfer by decreasing the distance between first period output targets. By decreasing the distance between first period output targets, she decreases the probability that she learns the agent's first period type. This reduces the high ability agent's incentive to mimic the low ability agent in the first period, and reduces the high ability agent's first period transfer.

Second, the principal decreases the low ability agent's first period transfer by increasing the distance between first period output targets. By increasing the distance between first period output targets, she increases the probability that she learns the agent's first period type. The low ability agent benefits when the principal learns his first period type; when the principal believes the agent has low ability in the first period, she designs the second period contract to induce more effort from the low ability agent. The higher is the low ability agent's effort in the second period, the higher is the high ability agent's information rent. Since the low ability agent receives no second period rent if his type remains low in the second period, he maximizes his expected second period rent by revealing his first period type to the principal. In this case, if his type switches between periods, he receives the highest possible second period rent.

Third, the principal increases her expected second period payoff by increasing the distance between first period output targets. When the principal believes she is fully informed about the agent's first period type, she strikes a better balance in the second period between inducing effort in the low ability agent and extracting rent from the high ability agent. If she believes that the agent's type is low in the first period, the second period contract induces more effort in the low ability type and thus leaves a higher rent for the high ability agent. When she believes that the agent's type is high in the first period, the second period contract calls for a lower effort in the low ability agent, which extracts rent from the high ability agent. Thus, the better is the principal's information about the agent's first period type, the more appropriately she can distort the low ability agent's second period effort.

From the perspective of the first period, therefore, her expected second period payoff is increasing in the probability that she learns the agent's first period type.

The combined effect of these three incentives determine how likely it is that the principal learns the agent's first period type. To give a frame of reference for the following discussion, we will say that the first period contract favors reducing the high ability agent's first period transfer if the optimal first period output targets, given in (2.30) and (2.31), lie within the commitment output targets given in (2.33) and (2.34). That is, if $\underline{y}_1^c \leq \underline{y}_1 \leq \overline{y}_1 \leq \overline{y}_1^c$, the principal learns less about the agent's first period type than she could by setting the first period output targets equal to the commitment optimum output targets.

Conversely, the first period contract favors reducing the low ability agent's first period transfer and increasing the principal's expected second period payoff if $\underline{y}_1 \leq \underline{y}_1^c \leq \overline{y}_1^c \leq \overline{y}_1$. In this case, the probability that the principal learns the agent's first period type is higher than if the principal set the first period output targets equal to the commitment optimum.

The following proposition shows that as long as either the high and low productivity agent do not differ too much in their ability, or the high productivity agent is not too unlikely ex-ante, the optimal first period contract favors reducing the upfront payment to the high ability agent.

Proposition 2.5. *If the high ability agent is not too much more productive than the low ability agent, or if the principal's prior belief that the agent is the high productivity type is not too low, then the overall impact of the first period contract is to reduce the distance between first period output targets, relative to the commitment optimum, for every level of positive correlation. Specifically, if* $\Theta^2 \geq 1/2$ *or* $\rho \geq 1/3$*, then*

$$\underline{y}_1^c \leq \underline{y}_1 \leq \overline{y}_1 \leq \overline{y}_1^c, \tag{2.47}$$

for every $\alpha \in [1/2,\, 1]$.

The interpretation of Proposition 2.5 is straightforward. First, consider the sufficient condition on the ratio of the agent's types, $\Theta^2 \geq 1/2$. Since $\Theta = \underline{\theta}/\overline{\theta}$, Proposition 2.5 states that as long as the low ability agent is at least 71 percent as productive as the high ability agent, the principal finds it optimal to design the first period contract to reduce the high productivity agent's first period transfer.

Recall that the principal reduces the high productivity agent's first period transfer by reducing the distance between first period output targets. Doing so reduces the difference in the low productivity agent's expected second period rent given that he targets $\overline{y}_1$, and his expected second period rent given that he targets $\underline{y}_1$. This in turn reduces his first period transfer.

The cost of bringing the output targets closer together is that the principal reduces the probability that she learns the agent's first period type. First, this increases the low productivity agent's first period transfer. However, Proposition 2.4 shows that the benefit of reducing the high productivity agent's first period transfer always outweighs the benefit of reducing the low productivity agent's first period transfer.

Second, when the principal is unsure of the agent's first period type, the low productivity agent exerts less effort in the second period than when the principal believes she is fully informed about the agent's first period type. The low productivity agent's effort distortion, however, is decreasing in Θ. Proposition 2.5 shows that as long as $\Theta^2 \geq 1/2$, the loss in the principal's expected second period payoff due to the increased probability of low effort from the low productivity agent is outweighed by the benefit of decreasing the first period transfer to the high productivity agent.

Similar reasoning explains the sufficient condition on the principal's beliefs that the agent is the high productivity type at the beginning of the interaction, $\rho \geq 1/3$. When the high productivity type is likely enough, the benefit of designing the first period contract to reduce the high productivity agent's first period transfer outweighs the second period reduction in the principal's payoff implied by the low productivity agent's reduced second period effort.

The following example illustrates how the optimal first period tradeoff between rent preservation and learning depends on the degree of positive correlation. In Figure 2.2, it is assumed that $\overline{\theta} = 1$, $\delta = 1$, and that noise is distributed uniformly on $[-1, 1]$. The principal's prior belief that the agent is the high productivity type is given by $\rho = .4$. The low ability agent is 80 percent as productive as the high ability agent (i.e. $\Theta = .8$). The horizontal axis measures the degree of positive correlation, and the vertical axis measures how much closer together the first period output targets are than the commitment output targets. One can clearly see that the difference between first period output targets is not monotone in α; $\overline{y}_1$ and $\underline{y}_1$ are closest together when $\alpha = .93$, and as α approaches one-half, they approach the commitment optimum targets.

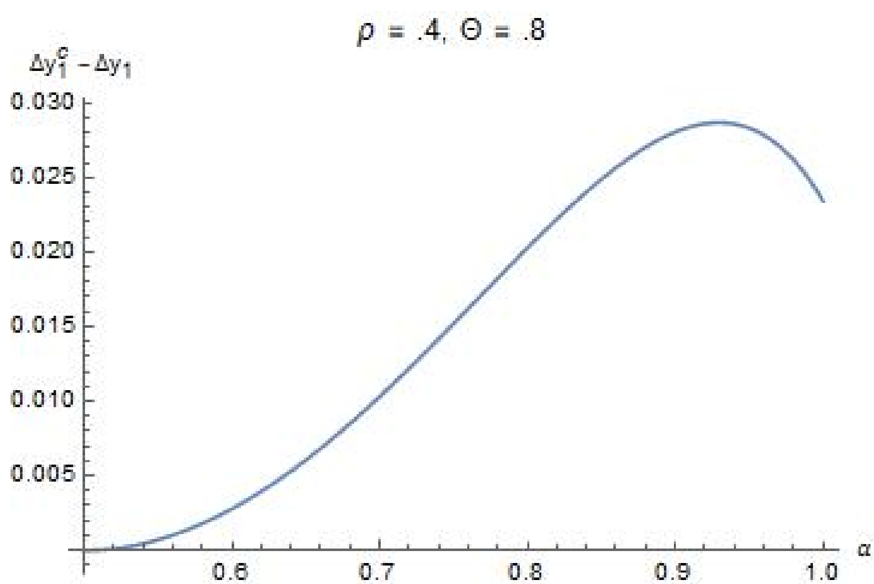

Figure 2.2: First period contract favors rent preservation for all α

When the sufficient conditions outlined in Proposition 2.5 do not hold, it is possible that the optimal first period output targets lie outside the commitment output targets. In the following example, assume once again that $\overline{\theta} = 1$, $\delta = 1$, and noise is distributed uniformly on $[-1, 1]$. Suppose, however, that the high productivity type is unlikely ex-ante ($\rho = .1$), and that the low ability agent is only 50 percent as productive as the high ability agent ($\Theta = .5$).

From Figure 2.3, one can see that for α between one-half and approximately three-fourths, the first period output targets lie outside the commitment output targets.[10] Thus, for smaller degrees of positive correlation, the optimal first period contract increases the probability that the principal learns the agent's fist period type relative to what she would learn under the commitment optimum. When the high productivity type is unlikely and the difference in the agent's productivity parameters is large, the benefit of reducing the low productivity agent's first period transfer, coupled with the benefit of better information in the second period, outweighs the benefit of reducing the high productivity agent's first period transfer. Still, this benefit only holds when the agent's type is weakly positively correlated.

2.6 Conclusion

This paper examines a repeated relationship between a principal and an agent. The agent produces output that the principal values. Two key features of this relationship are that the agent has imperfect control over output (output is stochastic), and the agent's private information may change over time (the agent's type is positively correlated).

The principal determines the probability that she learns the agent's first period type via

[10] That is, $\underline{y}_1 < \underline{y}_1^c < \overline{y}_1^c < \overline{y}_1$.

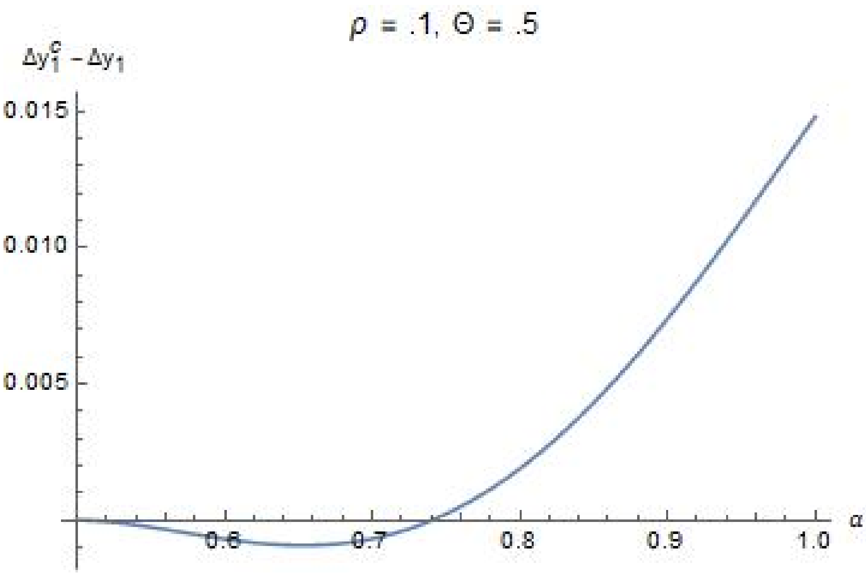

Figure 2.3: First period contract favors learning for some α

her choice of first period output targets. As long as the high productivity type is not too unlikely ex-ante, or as long as the difference in the agent's ability levels is not too large, the optimal first period contract favors reducing the high productivity firm's first period transfer. This is achieved by reducing the distance between first period output targets, relative to the commitment optimum.

The low productivity agent's first period transfer depends on the first period output targets; this feature does not arise in stochastic contracting models in which the agent's type is fixed over time. The low productivity agent's expected second period rent is increasing in the probability that the principal learns the agent's first period type. Therefore, to decrease the low productivity firm's first period transfer, the principal increases the distance between the first period output targets.

Both the high and low productivity agent's expected second period rents depend on the first period contract. Therefore, the question arises whether the principal reduces the prior-weighted sum of the high and low productivity agent's first period transfers by increasing or decreasing the distance between first period cost targets. Regardless of the principal's prior beliefs about the agent's type, this total expected transfer (prior-weighted sum) is reduced by decreasing the distance between output targets.

Chapter 3

Task Assignment Under Moral Hazard, with Effort-Dependent Human Capital and Outcome-Dependent Outside Options

3.1 Introduction

Firms want to develop their employees' skills. They want them to be better decision makers, better marketers, better at managing and motivating subordinates, and better product developers. One common way to develop skills is through training. Another potential avenue for developing human capital is learning by doing; specifically, the firm can assign workers to complete tasks, or implement projects. If the experience of implementing the project endows the worker with skills, this practice will benefit him and the firm in the future. The next time he undertakes the same task or one that is closely related to it, he will be better at it than he is today.

The above method of human capital acquisition has been discussed extensively, beginning with the work of Becker (1962). This paper examines a similar framework in which a principal is interested in developing her employees' human capital, but with the added supposition that the amount of human capital the agent develops depends positively on the amount of effort that he exerts on the task. If effort is costly to the agent, then the principal will have to induce the agent to exert effort in order to develop more than the minimum level of human capital.

Making this assumption allows us to examine human capital acquisition within the framework of the moral hazard literature. Following Mirrlees (1976) and Holmstrom (1979), when effort is unobservable, the principal will have to tie compensation to an observable and verifiable outcome in order to induce the desired level of effort from the agent.

In the absence of an outside labor market, nothing substantial changes between the classical moral hazard problem and the model augmented with human capital acquisition. If human capital confers some benefit to the principal, then the model augmented with human capital acquisition only increases the cost of effort at which the principal is indifferent between high and low effort.

However, the existence of an outside labor market may alter the agent's disutility of effort. If outside firms view a successful project as a signal that the employee worked hard, and developed valuable human capital, they may try to bid the agent away from his current firm when he has a success. Since exerting effort makes it more likely that the project will be successful, the increase in the expected value of the outside option from exerting high effort may outweigh the agent's increased cost of effort. If this is the case, then ex ante he prefers high effort to low effort, and we say the agent is self-motivated.

The introduction of the outside labor market borrows from the literature that discusses the theory of wage and promotion dynamics inside firms. Specifically, it is most closely related to Waldman (1984). Waldman considers an asymmetric learning model in which the agent's ability is initially unknown to the entire economy. The principal can assign the worker to one of two jobs: one that depends on his ability, and one that does not.

In equilibrium, the principal assigns all workers to the ability-independent job in the first period. After the first period production process, the principal learns the agent's ability perfectly. The outside market does not, but does observe the first period firm's decision to promote the agent to the ability-dependent job or keep him in the same job. The outside market uses this signal to update its beliefs about the agent's ability.

Due to the outside market's bidding behavior, which is influenced by the asymmetric information between the principal and the outside market, the principal promotes an inefficiently small number of agent-types in the second period. The agent must be productive enough in the new, ability dependent job, to offset the higher wage that the principal must pay him in order to keep him from leaving to work for the outside labor market.

In this paper, the principal and the outside market are not interested in learning about ability. They are instead interested in whether the agent has developed human capital. Like in Waldman (1984), the agent's first period employer has an informational advantage over the outside market. The principal observes the agent's effort, and so knows whether the agent developed human capital. The outside market is again left to infer the agent's level of human capital from a signal; in this paper, the signal is the success or failure of the project.

Unlike Waldman (1984), however, the market in this paper does not update expectations about the agent's effort choice using the project outcome. The market sends an outside offer if and only if the agent is successful. One can interpret this as the market over-valuing success; in this sense, success is costly to the first period employer. The firm wants to keep workers who exert effort and develop human capital, but they do so only if his benefit to the firm outweighs the increased cost of matching the outside offer.

In Waldman (1984), the principal's important strategic decision is whether to promote the agent or not after the first period is over. In this paper, the principal's strategic decision comes before the first period begins. She must decide whether the human capital that the agent will develop by exerting effort is valuable enough to offset the increased wage she would have to pay him, were he successful.

Consider the following real world example to motivate this paper. First, consider an assistant district attorney who develops litigation skills by taking cases to trial. The district attorney's office certainly values their employee's litigation skills, but should value their prosecutorial discretion more. That is, the district attorney's office wants a case to go to trial (or plea agreement) if the evidence warrants it. The assistant district attorney's chances of employment with a criminal defense firm, however, may be increasing in the number of visible successes that he has in court. Thus, there may arise cases in which the evidence against the defendant does not warrant a charge from the district attorney's point of view, but the assistant district attorney may nevertheless feel that the chances of a conviction are high.

Throughout the analysis, we restrict attention to the case in which the agent is self-motivated. Because success is costly to the principal, we examine the circumstances under which the principal prefers to induce low effort, when the agent is self-motivated. We say the contracting problem suffers from countervailing incentives when both of these conditions are met.

3.2 Model

Consider a two-period model, with one principal (she) and one agent (he). Both the principal and the agent are risk neutral, but the agent faces limited liability. In the first period, the principal decides whether to delegate completion of a project to the agent. If the agent exerts high (low) effort, the project succeeds with probability p_H (p_L), and fails with probability $1 - p_H$ ($1 - p_L$). When the project is successful, the principal realizes a profit π_S, and with failure, π_F. We assume $\pi_S \geq \pi_F > 0$. The project's outcome is observable and verifiable.

The agent incurs an effort cost $c(e_H) = \psi$ if he exerts effort, where $\psi \in [0, \infty)$. There is no cost of low effort ($c(e_L) = 0$). We assume that the principal can observe $e \in \{e_H, e_L\}$, but effort is not verifiable, and so contracts must be written on the project's outcome alone.

We augment the model with human capital acquisition and an outside labor market. If the agent exerts high (low) effort when implementing the project, he gains human capital $v(e_H) = v$ ($v(e_L) = 0$). We can think about this as learning by doing, with the added assumption that the amount of human capital developed depends on how much effort the agent exerts while undertaking the project. The key point is that if the agent shirks in Period 1, he does not improve or acquire any skills that the principal values, whereas if he exerts high effort, he develops these skills whether the project succeeds or fails.

The principal and the outside labor market value the skills that the agent develops by exerting effort. However, they may have different valuations for his newly developed skills. At the end of the first period, the outside market observes the outcome of the project. If the project is a success, the outside market sends an offer of α to the agent. The labor market is a "black box" in the sense that the agent gets an outside offer at the beginning of the second period when he is successful, regardless of whether he actually exerted effort and gained human capital. Similarly, he gets no outside offer when he exerts high effort but fails.

We allow $\alpha \in [0, \infty)$. When $0 \leq \alpha < v$, outside firms value his human capital less than the inside firm, in which case the human capital developed is more firm specific. When $\alpha = 0$, we have the classic case of purely firm specific human capital. However, we allow $\alpha \geq v$ to capture the idea that the skills a worker develops may be more useful or productive outside his current firm.

To see an example in which the principal and an outside labor market may have different valuations for the human capital developed on a task, recall the assistant district attorney who develops litigation experience by taking cases to trial. The district attorney only values the assistant's litigation experience on a given case if the evidence warrants going to trial. A private criminal defense firm, however, may value the litigation experience regardless of the merit of the case.

The principal observes the agent's outside offer, and decides whether to match it or let him leave. The principal only obtains the human capital payoff if she retains the agent, and the agent exerted effort in the first period. If she lets the agent leave, she faces a cost c of replacing the agent, whether or not the agent exerted effort in the first period.

The outside offer has two important effects on the principal's decision making. First, when α is large enough, the agent is too expensive to keep, regardless of first period effort. Second, the outside offer will create the possibility that the agent is "self motivated." We will use this phrase to mean that in the absence of a wage, the agent prefers to exert high effort rather than low effort (Under the usual contracting problem under moral hazard, the "status-quo" level of effort is low effort, since effort is costly). When α is large enough, the agent prefers to exert effort, and increase his probability of getting an outside offer, even when effort is costly. Notice that when the agent's cost of effort is small, the outside offer does not have to be large in order for the agent to be self motivated.

Self motivation will plays an important role in the analysis to follow. We concentrate on the contracting problem between the principal and the agent, when the agent is self motivated. We characterize parameter restrictions that ensure the existence of countervailing incentives, which occur when when the agent is self motivated, but profit maximization dictates that the principal induce low effort.

The timing of the game is as follows: In Period 1, the principal offers the agent a contract $w = (w_S, w_F)$. The agent chooses his effort, and at the end of Period 1, the principal realizes project payoff π_o, and pays the agent his outcome contingent wage, w_o, where $o \in \{S, F\}$.

At the end of Period 1, outside firms observe whether the agent successfully implemented the project. If he did, then he gets a wage offer α at the beginning of Period 2. The principal observes α, and has the opportunity to match the outside wage offer. If she chooses to match, then at the end of Period 2, she realizes the second period human capital payoff, v, less the outside offer. Her second period payoff is then $U = v - \alpha$. If she chooses not to match, she does not get v, and incurs cost $c > 0$ of replacing the agent.

The principal's payoff depends on whether she matches the outside offer and whether the agent develops human capital:

$$U = \pi_o - w_o + v(e) - \alpha, \tag{3.1}$$

where $v(e) = v$ if the agent exerted high effort, and zero otherwise. Likewise, if the project fails, $\alpha = 0$. The agent's payoff is as follows:

$$u = w_o - c(e) + \alpha, \tag{3.2}$$

where $c(e) = \psi$ if the agent exerts effort, and zero otherwise. Again, if the project fails, $\alpha = 0$. Figure 3.1 illustrates the timing of the game, the principal's second period strategy choices, and the principal and agent's outcome contingent payoffs.

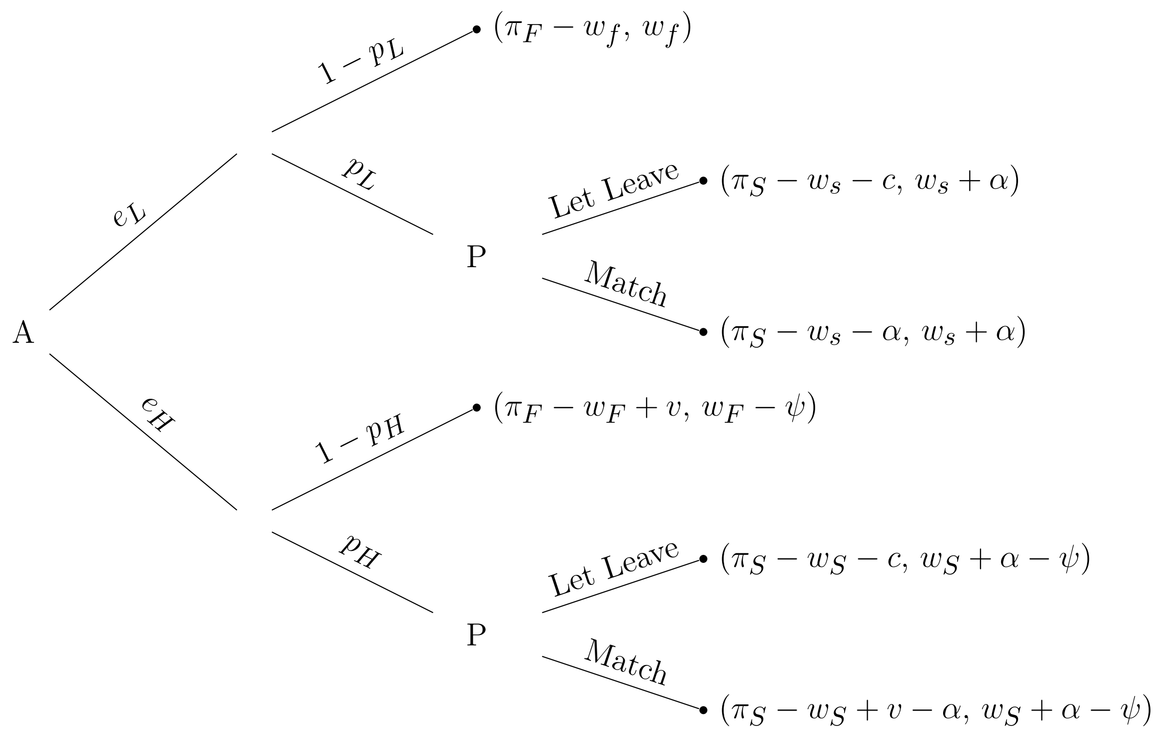

Figure 3.1: Game tree and payoffs (U_e, u_e), $e \in \{L, H\}$

When the agent exerts low effort he does not develop human capital. The project succeeds with probability p_L, and fails with probability $1 - p_L$. The principal's payoff from low effort, when she chooses "Let Leave" and "Match" in the second period, are

$$U_L^L = p_L(\pi_S - w_s - c) + (1 - p_L)(\pi_F - w_f) \tag{3.3}$$

and

$$U_M^L = p_L(\pi_S - w_s - \alpha) + (1 - p_L)(\pi_F - w_f), \tag{3.4}$$

respectively. If she matches, she must pay the value of the outside offer to the agent to retain him. If she lets him leave, she incurs cost c of replacing the agent. Comparing (3.3) and (3.4), we can see that when the agent exerts low effort, the principal will match an outside offer when

$$\alpha < c. \tag{3.5}$$

When the agent exerts effort, he develops human capital. The project succeeds with probability p_H, and fails with probability $1 - p_H$. The principal's payoff from high effort, when she chooses "Let Leave" and "Match" in the second period are

$$U_L^H = p_H(\pi_S - w_S - c) + (1 - p_H)(\pi_F - w_F + v) \tag{3.6}$$

and

$$U_M^H = p_H(\pi_S - w_S + v - \alpha) + (1 - p_H)(\pi_F - w_F + v), \tag{3.7}$$

respectively. Comparing (3.6) and (3.7), we can see that when the agent exerts effort, the principal will play "Match" if and only if

$$\alpha < c + v. \tag{3.8}$$

3.3 Analysis

The principal's decision of which effort level to implement depends on the size of the outside offer that the agent receives when he is successful, the agent's cost of exerting high effort, and the difference in expected project payoffs when the agent exerts high and low effort. To implement high effort, the principal must satisfy the following incentive and participation constraints:

$$p_H w_S + (1 - p_H) w_F + p_H \alpha - \psi \geq p_L w_S + (1 - p_L) w_F + p_L \alpha \tag{3.9}$$

$$p_H w_S + (1 - p_H) w_F + p_H \alpha - \psi \geq 0. \tag{3.10}$$

We derive the condition for self-motivation from the agent's incentive constraint, (3.9). When the agent is self-motivated, (3.9) is satisfied in the absence of a wage (i.e. when $w_S = w_F = 0$):

$$p_H \alpha - \psi \geq p_L \alpha. \tag{3.11}$$

From (3.11) follows the definition of a self motivated agent:

Definition 3.1. *The agent is self motivated when the outside offer is large enough that the agent prefers to exert high effort in the absence of a wage. That is, when*

$$\Rightarrow \alpha \geq \frac{\psi}{\Delta p}. \tag{3.12}$$

This is the opposite of the canonical contracting problem under moral hazard, in which effort is costly and the agent must be offered an incentive wage to exert effort. In such a case, the principal rewards success and punishes failure when she wants the agent to exert effort, since success is more likely when the agent exerts effort. If she wants the agent to shirk, she offers a wage that is the same whether the project succeeds or fails, and just satisfies the agent's participation constraint. When the agent is self-motivated, she must punish success and reward failure if she wants the agent to shirk, while she can offer a flat wage if she wants the agent to exert effort. The agent's incentive and participation constraints to implement low effort are as follows:

$$p_L w_S + (1 - p_L) w_F + p_L \alpha \geq p_H w_S + (1 - p_H) w_F + p_H \alpha - \psi \tag{3.13}$$

$$p_L w_S + (1 - p_L) w_F + p_L \alpha \geq 0. \tag{3.14}$$

If both (3.9) and (3.10) hold with equality, the optimal (unlimited liability) wage schedule to induce high effort is given by[1]

$$w_S = \frac{(1 - p_L)}{\Delta p} \psi - \alpha, \tag{3.15}$$

$$w_F = \frac{-p_L}{\Delta p} \psi. \tag{3.16}$$

Since the agent has limited liability, however, the principal must pay the agent a non-negative wage. When the agent is self motivated, (3.15) and (3.16) are negative. Therefore, the principal implements the following wage schedule to induce high effort:

$$w_S^* = w_F^* = 0. \tag{3.17}$$

Substituting these wages into (3.7) and (3.6), the principal's equilibrium profit when the

[1] This is also the optimal unlimited liability wage schedule to induce low effort.

agent exerts effort and she matches the outside offer is

$$U_M^H = p_H(\pi_S + v - \alpha) + (1 - p_H)(\pi_F + v), \tag{3.18}$$

and when she lets the agent leave,

$$U_L^H = p_H(\pi_S - c) + (1 - p_H)(\pi_F + v). \tag{3.19}$$

To induce low effort, the best the principal can do is set the wage the agent receives upon a success equal to zero, and make the agent's incentive constraint (3.13) bind. This results in the following wage schedule:

$$\begin{aligned} w_s &= 0, \\ w_f &= \alpha - \frac{\psi}{\Delta p}. \end{aligned} \tag{3.20}$$

Substituting these wages into (3.4) and (3.3) the equilibrium profit for the principal when the agent exerts low effort and she matches the outside offer is

$$U_M^L = p_L(\pi_S - \alpha) + (1 - p_L)\left(\pi_F - \alpha + \frac{\psi}{\Delta p}\right), \tag{3.21}$$

and when she lets the agent walk,

$$U_L^L = p_L(\pi_S - c) + (1 - p_L)\left(\pi_F - \alpha + \frac{\psi}{\Delta p}\right). \tag{3.22}$$

In what follows, we break the analysis up based on the size of α. In Region 1 of Figure 3.2, α is small enough so that principal matches an outside offer whether the agent exerts high or low effort; that is, Region 1 is defined by $\alpha < c$. In Region 2, the principal matches an outside offer if the agent exerts high effort, but not if the agent exerts low effort; that is, Region 2 is defined by $\alpha \in [c, c+v]$. In Region 3, the principal lets the agent walk, regardless of first period effort; that is, Region 3 is defined by $\alpha > c + v$.

We study the principal's decision making in each region. Specifically, we are interested in situations in which the principal induces a self-motivated agent to shirk. We call this occurrence countervailing incentives:

Definition 3.2. *The contracting problem between the principal and the agent suffers from **countervailing incentives** when the agent is self motivated, but the principal's profit is maximized by inducing the agent to shirk.*

A necessary condition for countervailing incentives to arise is that the agent is self motivated. Therefore, in the analysis to follow, we restrict attention in each Region to outside offers $\alpha \geq \frac{\psi}{\Delta p}$. In Figure 3.2, the function $\alpha(\psi)$ gives, for every cost of effort ψ, the outside offer that just makes the agent self-motivated. Therefore, for $\alpha < \alpha(\psi)$, the agent with cost of effort ψ must be incentivized to exert high effort. For $\alpha \geq \alpha(\psi)$, the agent with cost ψ is self motivated, and the principal must reward failure to induce the agent to exert low effort. In each Region, the analysis to follow is restricted to outside offers large enough that the agent is self motivated (that is, outside offers that lie above $\alpha(\psi)$.

3.3.1 Existence of countervailing incentives, Region 1

In this sub-section, characterize the existence of countervailing incentives in Region 1. Recall that Region 1 captures all outside offers that are less than the cost of replacing the agent, $\alpha < c$.

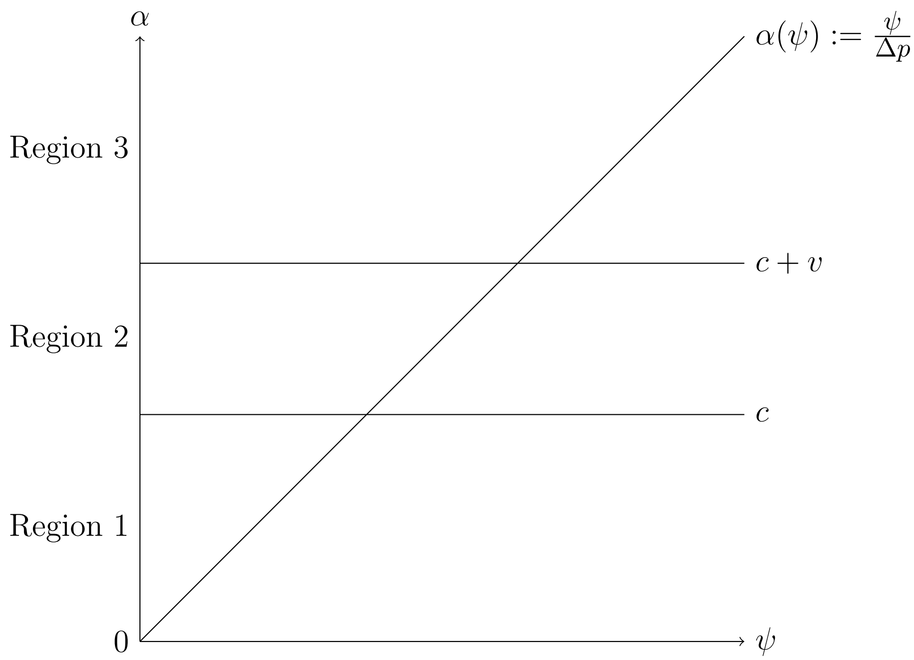

Figure 3.2: The agent is self motivated for $\alpha \geq \alpha(\psi)$.

Given that $\alpha < c$, the largest cost of effort that the agent can possess and still face countervailing incentives in Region 1 is

$$\psi_c := c \cdot \Delta p. \tag{3.23}$$

Note that ψ_c is the cost of effort at which $\alpha(\psi) = c$. If the agent has cost of effort ψ_c, the smallest outside offer which makes him self-motivated is $\alpha = c$. To begin the analysis of countervailing incentives, consider Table 3.1, which gives the principal's expected revenues (row 1) and expected costs (row 2) when the agent exerts high effort.

	Period 1	Period 2
E[R]	$p_H \pi_S + (1 - p_H)\pi_F$	v
E[C]	0	$p_H \alpha$

Table 3.1: Expected revenues and expected costs when $e = e_H$

In Period 1, the project either succeeds or fails. Because the agent is self motivated, the principal does not have to pay the agent to exert high effort (see (3.17)). Therefore, expected costs in the first period are equal to zero. In Period 2, the principal obtains the benefit of the human capital payoff, v, because the agent exerted high effort. With probability p_H, the agent receives an outside offer α. Since this outside offer is less than the cost of replacing the agent, the principal decides to match the outside offer and retain the agent. Thus, expected costs in Period 2 are $p_H\alpha$.

	Period 1	Period 2
E[R]	$p_L\pi_S + (1-p_L)\pi_F$	0
E[C]	$(1-p_L)(\alpha - \frac{\psi}{\Delta p})$	$p_L\alpha$

Table 3.2: Expected revenues and expected costs when $e = e_L$

When the agent exerts low effort (see Table 3.2), both first and second period expected revenues decrease. First period expected revenue decreases because the project is less likely to be successful, and second period expected revenue decreases because the agent does not develop human capital when he exerts low effort. Expected costs increase in the first period and decrease in the second period, relative to when the agent exerts high effort. The principal has to reward failure in order to incentivize low effort; therefore, when the project fails, the principal pays the agent $w_f = \alpha - \psi/\Delta p$. However, she is less likely to have to match an outside offer, which decreases expected second period costs from $p_H\alpha$ to $p_L\alpha$.

Expected revenues are lower when the agent exerts low effort than when the agent exerts high effort. Therefore, it is necessary for expected costs under low effort to be lower than expected costs under high effort for the principal's expected profits under low effort to be higher than her expected profits under high effort. The left hand side of the first inequality in (3.24) is the principal's expected cost of inducing low effort, and the right hand side is the expected cost when the agent exerts high effort:

$$p_L\alpha + (1 - p_L)(\alpha - \frac{\psi}{\Delta p}) < p_H\alpha$$
$$\Rightarrow \alpha < \frac{1 - p_L}{1 - p_H}\frac{\psi}{\Delta p}. \tag{3.24}$$

Since $p_H > p_L$, we know that the probability of failure conditional on low effort is larger than the probability of failure conditional on high effort. Thus, the cost of inducing low effort is lower than the cost of inducing high effort when the agent is "not too self motivated," that is, when

$$\alpha \in \left[\frac{\psi}{\Delta p}, \frac{1 - p_L}{1 - p_H}\frac{\psi}{\Delta p}\right]. \tag{3.25}$$

Recall that countervailing incentives exist when the agent is self-motivated and the principal's expected profits from low effort are higher than her expected profits from low effort. Therefore, countervailing incentives will exist for some sub-set of $\left[\frac{\psi}{\Delta p}, \frac{1-p_L}{1-p_H}\frac{\psi}{\Delta p}\right]$.

From (3.18) and (3.21), the principal's expected profits under low effort are higher than her expected profits under high effort when

$$U_M^H < U_M^L \Rightarrow \alpha < \frac{1 - p_L}{1 - p_H}\frac{\psi}{\Delta p} - \frac{\Delta p \Delta \pi + v}{1 - p_H} =: \alpha_1(\psi). \tag{3.26}$$

We refer to $\alpha_1(\psi)$ as the principal's Region 1 decision rule. This decision rule reflects the principal's profit maximizing allocation of effort. For every value of ψ, it gives the outside offer that makes the principal indifferent between high and low effort. For $\alpha > \alpha_1(\psi)$, the principal induces high effort, and for $\alpha < \alpha_1(\psi)$, the principal induces low effort.

For countervailing incentives exist in Region 1, it must be the case that the principal prefers low effort, but the agent is self motivated. Therefore, it must be the case that

$$\begin{aligned} &\frac{\psi}{\Delta p} < \frac{1-p_L}{1-p_H}\frac{\psi}{\Delta p} - \frac{\Delta p \Delta \pi + v}{1-p_H} \\ &\Rightarrow \psi > \Delta p \Delta \pi + v =: \hat{\psi}. \end{aligned} \tag{3.27}$$

The above condition is necessary for countervailing incentives to exist in Region 1, but it is not sufficient. Cost of effort $\hat{\psi} := \Delta p \Delta \pi + v$ is the smallest cost of effort for which it is possible that an outside offer makes the principal indifferent between high and low effort profits, and the agent self motivated. Put differently, for every cost of effort less than $\hat{\psi}$, whenever the outside offer is low enough to make low effort profits optimal, the agent is not self motivated.

Consider further the meaning of the necessary condition in (3.27). When the agent exerts high effort, expected project revenues increase by $\Delta p \Delta \pi$. Further, the agent develops human capital, which is valuable to the principal. Therefore, what (3.27) shows is that the agent's cost of effort must be high large enough to outweigh this increase in expected revenues. This reasoning also explains why (3.27) is not sufficient for countervailing incentives to exist in Region 1. As discussed above, the largest cost of effort in Region 1 is ψ_c. Therefore, for countervailing incentives to exist in Region 1, it must be the case that $\hat{\psi} < \psi_c$. The following proposition provides a sufficient condition for this to be the case.

Proposition 3.1. *If the cost associated with replacing the agent is large enough, then countervailing incentives exist in Region 1. That is, if*

$$c \geq \frac{\Delta p \Delta \pi + v}{\Delta p}, \tag{3.28}$$

then countervailing incentives exist in Region 1.

Proof. Suppose $c \geq \frac{\Delta p \Delta \pi + v}{\Delta p}$. Then

$$\Delta pc \geq \Delta p \Delta \pi + v$$
$$\Rightarrow \psi_c \geq \hat{\psi}. \tag{3.29}$$

Notice that $\hat{\psi}$ represents the intersection between $\alpha_1(\psi)$ and the principal's Region 1 decision rule. Since $\hat{\psi} < \psi_c$, this intersection occurs in Region 1. Since this intersection occurs in Region 1, for $\psi \in [\hat{\psi}, \psi_c]$ there exist outside offers $\alpha < c$ for which the agent is self motivated and the principal induces low effort (see Figure 3.3). □

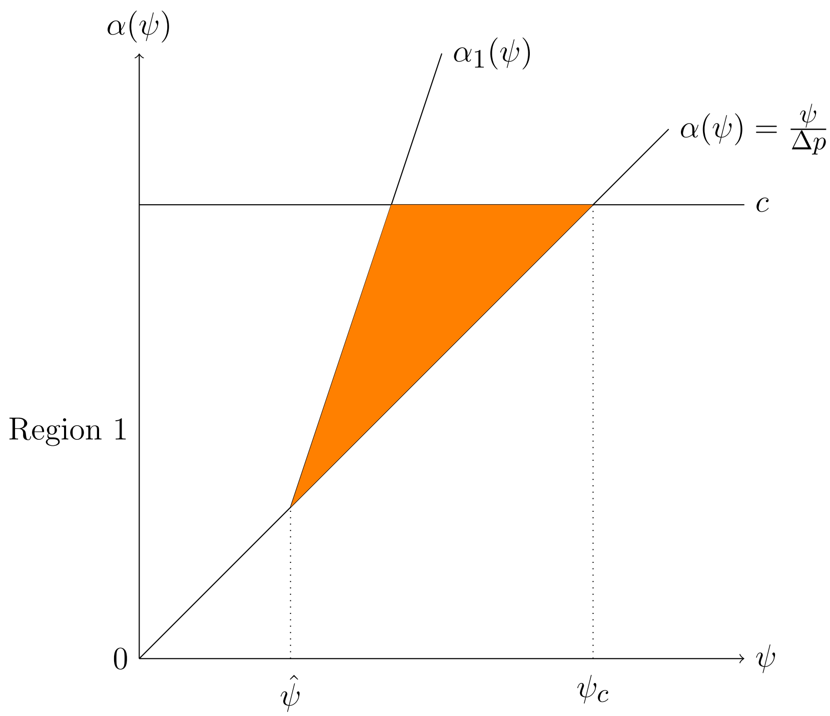

Figure 3.3: Countervailing incentives, Region 1

Proposition 3.1 shows that the cost of replacing the agent must be large enough for countervailing incentives to occur in Region 1. This is true even though the principal never incurs the cost of replacing the agent for outside offer $\alpha < c$. The larger is c, however, the larger are the outside offers contained in Region 1. When c is large enough, the cost of effort that makes the principal indifferent between high and low effort, $\hat{\psi}$, is less than the largest cost of effort in Region 1, ψ_c.

This is illustrated in Figure 3.3. The principal induces the agent to shirk in the shaded region. In this shaded area, the agent with cost of effort $\psi \in [\hat{\psi}, \psi_c]$ is self motivated, because the outside offer is larger than $\alpha(\psi)$. Further, the principal's profit maximizing effort choice is low effort because $\alpha < \alpha_1(\psi)$.

When countervailing incentives exist, it is not because high effort wage payments are prohibitively expensive; the agent is self motivated, so it is costless for the principal to induce high effort. Recall that countervailing incentives exist only if the agent is not too self motivated (see (3.25)). When the agent is not too self motivated, the cost of inducing low effort is small. When the cost of replacing the agent is large, the outside offers in Region 1 are large. Therefore, countervailing incentives exist in Region 1 when the outside offer is close to the cost of replacing the agent (i.e. when α is close to c) and the cost of inducing low effort is small (i.e. the agent is not too self motivated).

3.3.2 Countervailing incentives, Regions 2 and 3

In Region 2, the principal matches the outside offer only if the agent exerted effort and developed human capital. If the agent was successful despite shirking, she lets the agent leave in the second period. The principal's profits from high and low effort are given by

(3.18) and (3.22), respectively.[2] Therefore, the principal induces low effort if

$$U_M^H < U_L^L \implies \alpha > \frac{1}{1 - p_H - p_L} \left[\frac{1 - p_L}{\Delta p} \psi - \hat{\psi} - p_L c \right] =: \alpha_2(\psi). \tag{3.30}$$

Above, $\alpha_2(\psi)$ represents the principal's Region 2 decision rule; the principal implements low effort when $\alpha \geq \alpha_2(\psi)$, and high effort when $\alpha < \alpha_2(\psi)$.

In Region 3, the outside offer is so large that the principal lets the agent leave, regardless of his first period effort choice. Therefore, the principal's profits from high and low effort are given by (3.19) and (3.22), respectively. The principal induces low effort if

$$U_L^H < U_L^L \implies \alpha < \frac{\psi}{\Delta p} + (\psi_c - \hat{\psi}) + p_H v =: \alpha_3(\psi). \tag{3.31}$$

Once again, we refer to $\alpha_3(\psi)$ as the principal's Region 3 decision rule. The principal implements low effort if $\alpha \leq \alpha_3(\psi)$, and implements high effort if $\alpha > \alpha_3(\psi)$.

We state without formal proof that if countervailing incentives exist in Region 1, they exist in Regions 2 and 3. The intuition is easy to see from the principal's decision rule in Region 3. Recall that countervailing incentives exist in Region 1 if $\psi_c > \hat{\psi}$. From (3.31), if $\psi_c > \hat{\psi}$, then $\alpha_3(\psi)$ lies above $\alpha(\psi)$. If $\alpha(\psi) < \alpha_3(\psi)$, then for outside offers $\alpha \in [\alpha(\psi), \alpha_3(\psi)]$, the agent is self motivated and the principal induces low effort. Therefore, if countervailing incentives exist in Region 1, they exist in Region 3.

The principal's behavior can be explained in a similar manner to her behavior in Region 1. The wage-cost of inducing high effort is zero, but on the chance that the agent is successful, the principal must pay cost c to replace the agent. When the agent is not too self motivated,

[2] It is assumed that the average probability of success is greater than one half. That is, $1 < p_H + p_L$. This ensures the decision rule in Region 2 is downward sloping.

the cost of inducing low effort is small. By inducing low effort, the principal reduces the probability that she has to replace the agent. Thus, when c is large enough, the principal benefits from inducing the agent that is not too self motivated to shirk.

In Region 2, the intuition is less clear. Notice that the principal's decision rule in Region 2 is negatively sloped, since $1 - p_H - p_L < 0$. The reason we study this case is because it creates a non-monotonicity in the agent's profit maximizing allocation of effort, which we examine in more detail below.

For the purpose of showing that countervailing incentives exist in Region 2, we can make an argument similar to the one for the existence of countervailing incentives in Region 1. If countervailing incentives exist in Region 1, we can show that the smallest cost of effort in Region 2 for which countervailing incentives can exist, which we will denote $\check{\psi}$, is less than the largest cost of effort in Region 2, which we denote ψ_{c+v}.

Notice that $\check{\psi}$ is the Region 2 equivalent of $\hat{\psi}$. For every cost of effort less that $\check{\psi}$, the principal will induce high effort in Region 2. Similarly, ψ_{c+v} is the Region 2 equivalent of ψ_c. When the agent has cost of effort ψ_{c+v}, the smallest outside offer which makes him self motivated is $\alpha = c + v$, which is the upper bound on Region 2. For $\psi \in (\check{\psi}, \psi_{c+v})$, countervailing incentives occur when

$$\alpha \in \Big(max\{\alpha_3(\psi),\, c\},\, c + v\Big). \tag{3.32}$$

The intuition is the same as in Regions 1 and 3. When the agent is not too self motivated in Region 2, the principal is better off inducing low effort.

3.3.3 Analysis of contracting for a fixed cost of effort

Now that we have characterized the existence of countervailing incentives in Regions 1, 2, and 3, and discussed the intuition for why the principal induces a self motivated agent to shirk, we examine the contracting game between the principal and the agent when the agent's cost of effort is $\bar{\psi} \in (\hat{\psi}, \tilde{\psi})$. [3] We restrict ourselves to these costs of effort because we are interested in the non-monotonicity in the principal's profit maximizing allocation of effort when $1 - p_H - p_L < 0$.[4]

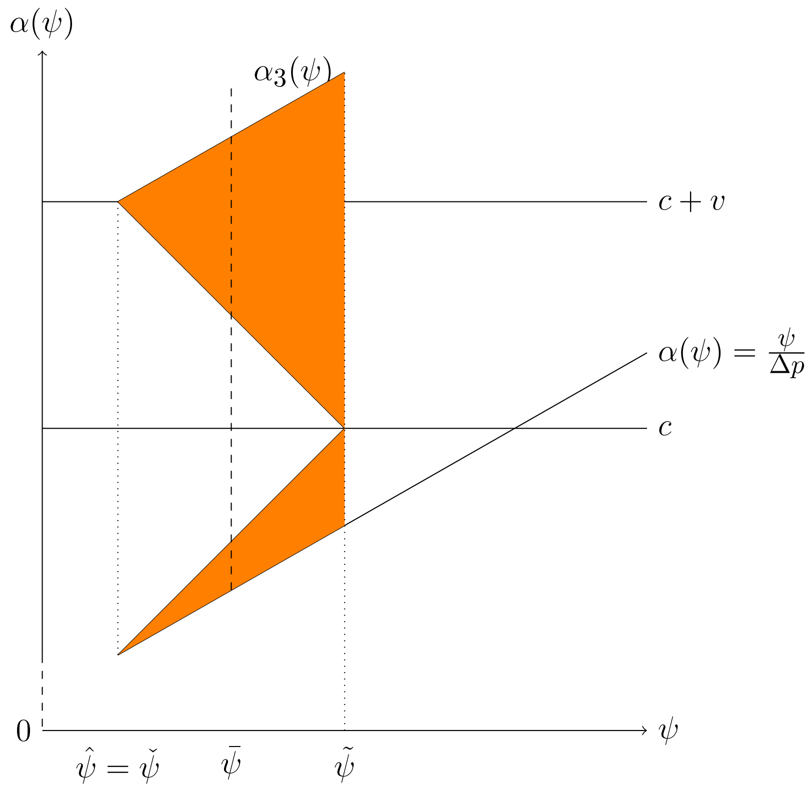

Figure 3.4: Countervailing incentives, $p_H + p_L > 1$

In the shaded areas of Figure 3.4, the principal induces the agent with cost of effort $\bar{\psi}$ to shirk, while in the non-shaded areas, he is allowed to exert effort. As the outside offer increases, the principal's profit maximizing choice of effort changes three times. We will

[3] $\tilde{\psi}$ is the cost of effort at which $\alpha_1(\psi)$ is equal to c

[4] For graphical simplicity, we assume that $\check{\psi} = \hat{\psi}$, however, they need not be equal.

examine why the optimal allocation of effort, and thus the decision of whether to allow human capital acquisition or not, is so sensitive to the size of the outside offer when $1-p_H-p_L<0$.

We briefly discussed the principal's profit maximizing effort choice when the outside offer is in Region 1. Here, we will re-visit that process for fixed cost of effort $\bar{\psi}$. We can show that the principal is actually better off when the agent fails, regardless of the first period effort choice. Recall that the principal's payoff, when the agent exerts high effort and fails, is

$$U_M^H|F = \pi_F + v. \tag{3.33}$$

When the agent exerts high effort and succeeds, her payoff is

$$U_M^H|S = \pi_S + v - \alpha \tag{3.34}$$

Notice that

$$U_M^H|F - U_M^H|S = \alpha - \Delta\pi, \tag{3.35}$$

which is positive since the agent is self motivated and $\bar{\psi} > \hat{\psi}$.

Similarly, the principal is better off when the agent fails after exerting low effort:

$$U_M^L|F - U_M^L|S = (\pi_F - (\alpha - \frac{\bar{\psi}}{\Delta p})) - (\pi_S - \alpha) = \frac{\bar{\psi}}{\Delta p} - \Delta\pi, \tag{3.36}$$

which is also positive, since $\bar{\psi} > \hat{\psi}$.

Further, we know that $U_M^H|S - U_M^L|S > 0$, since the only difference between the two payoffs is that the agent develops valuable human capital when he exerts effort, and does not develop human capital when he shirks. We can also show that $U_M^H|F - U_M^L|F > 0$, since the principal must pay to induce low effort, and gets no human capital payoff. Lastly, we can determine that $U_M^L|F - U_M^H|S > 0$, since $\bar{\psi} > \hat{\psi}$.

This last ordering may seem counter intuitive; the principal is better off when she pays the agent to exert low effort, gain no human capital, and the project fails, than she is when the agent exerts high effort "for free," gains human capital, and the project succeeds. This has to do with the structure of the wage to induce low effort, and the fact that the principal must match the outside offer once the agent succeeds.

We are left with the following general ranking of the principal's high and low effort profits in Region 1, conditional on outcome:

$$U_M^L|S < U_M^H|S < U_M^L|F < U_M^H|F. \tag{3.37}$$

The principal is best off, regardless of first period effort choice, when the agent fails. Given our assumptions on p_H and p_L, we can now explain the principal's decision making in Region 1.

We can more clearly see the reasoning if we rearrange the principal's decision rule. Recall that in Region 1, the principal's high and low effort payoffs are given by (3.18) and (3.21), respectively. To determine why the principal induces low effort in an agent who is just self motivated, and switches to high effort as the outside offer passes $\alpha_1(\cdot)$, consider the expression:

$$U_M^L - U_M^H = \Big[\Delta p(\pi_F - (\pi_S - \alpha))\Big] - \Big[(1 - p_L)(\alpha - \frac{\bar{\psi}}{\Delta p}) + v\Big]. \tag{3.38}$$

The first term in brackets in (3.38) is the principal's benefit of inducing low effort, relative to high effort. We showed in (3.37) that the principal is better off when the agent fails than when he succeeds, regardless of effort choice. This is illustrated again here; under low effort, the project fails with probability $(1 - p_L)$, and under high effort, it fails with probability $(1 - p_H)$. Then the probability of getting the project payoff from failure is increased by $p_H - p_L$ when the principal induces low effort, and the probability of getting the payoff from success, less the cost of matching the outside offer, is decreased by $p_H - p_L$. This benefits the principal because when the cost of effort is greater than $\hat{\psi}$, if the outside offer is large enough to make the agent self-motivated, the project payoff from failure is higher than the project payoff from success, less the cost of matching the outside offer.

The second term in brackets in (3.38) is the principal's cost of inducing low effort. The first term is the wage cost of inducing low effort, and the second term is the opportunity cost of not inducing human capital acquisition. Notice that the benefit of inducing low effort, relative to high effort, is increasing in α at rate $p_H - p_L$, while the cost of inducing low effort is increasing at rate $1 - p_L > p_H - p_L$. Thus, eventually, the cost of inducing low effort outweighs the benefit. This occurs exactly when $\alpha = \alpha_1(\bar{\psi})$.

As α increases past $\alpha_1(\bar{\psi})$ in Region 1, the principal maintains high effort. As α crosses c, the principal will no longer match the outside offer if the agent exerted low effort. Thus, the principal's payoff is still given by (3.18) if the agent exerted effort in the first period, but is now given by (3.22) if the agent shirked.

We analyze Region 2 similarly to how we analyzed Region 1. We can obtain a similar general ordering of payoffs in Region 2, conditional on outcome, as we did in Region 1. The conditional payoffs $U_M^H|S$ and $U_M^H|F$ are unchanged, so we still have $U_M^H|F > U_M^H|S$, for the same reasons as before. The principal's conditional payoffs from low effort are now given by:

$$U_L^L|S = \pi_S - c \tag{3.39}$$

and

$$U_L^L|F = \pi_F - (\alpha - \frac{\bar{\psi}}{\Delta p}). \tag{3.40}$$

The principal's payoff from low effort, conditional on failure, is still higher than her payoff from high effort, conditional on success:

$$U_L^L|F - U_L^L|S = \frac{\bar{\psi}}{\Delta p} - (\alpha - c) - (\Delta\pi) > 0. \tag{3.41}$$

This is because, in Region 2, $\alpha - c < c + v - c = v$, and $\bar{\psi} > \hat{\psi}$. Further, we know that

$$U_M^H|S - U_L^L|S = c + v - \alpha > 0, \tag{3.42}$$

since $\alpha < c + v$. Recall that in Region 1, $U_M^H|S - U_M^L|S = v$ does not depend on the outside offer. In Region 2, profits from success in the high effort state approach profits from success in the low effort state as α approaches $c + v$. This has an important effect on the rate at which the benefit from inducing low effort increases in α.

Additionally, since $U_L^L|F$ is the same expression as $U_M^L|F$, we know that $U_M^H|F > U_L^L|F$ and $U_L^L|F > U_M^H|S$, for the same reasons as in Region 1. Then we are left with a general ordering similar to the one we had in Region 1:

$$U_L^L|S < U_M^H|S < U_L^L|F < U_M^H|F. \tag{3.43}$$

Again, the principal is best off, regardless of effort choice, when the agent fails. We can now explain the principal's decision making in Region 2.

We can rearrange the principal's decision rule to see how the principal's benefits and costs from inducing low effort change as α increases inside Region 2:

$$U_L^L - U_M^H = \Big[\Delta p \pi_F - (p_H(\pi_S - \alpha) - p_L(\pi_S - c))\Big] - \big[(1-p_L)(\alpha - \frac{\bar{\psi}}{\Delta p}) + v\big]. \quad (3.44)$$

The first term in brackets in (3.44) is the principal's benefit from inducing low effort relative to high effort. Again, the fact that the principal is better off when the agent fails is reflected here. The principal's probability of getting the project payoff from failure increases by $p_H - p_L$ when she induces low effort instead of high effort. The principal's probability of getting the payoff from success decreases, but so does her cost of matching the outside offer. When the agent exerts low effort and succeeds, the principal lets him leave, and pays $c < \alpha$ to replace him.

The second term in brackets is the principal's cost of inducing low effort, relative to high effort, and is the same expression as in Region 1. From (3.44), we see that the principal's benefit from inducing low effort relative to high effort is increasing at rate p_H in α, while her cost is increasing at rate $1 - p_L$. Since $1 - p_H - p_L < 0 \Rightarrow p_H > 1 - p_L$, her benefit of inducing low effort relative to high effort is increasing more rapidly in α than her cost. Once $\alpha > \alpha_2(\psi)$, she will induce low effort.

As α increases, the principal's payoff from high effort, conditional on success, approaches the principal's payoff from low effort conditional on success. Since $p_H - p_L$ is large enough that countervailing incentives exist, the principal is better off shifting away from high effort, even though her best possible payoff is $U_M^H|F$. This payoff is sufficiently unlikely that the principal prefers low effort as α increases.

As α enters Region 3, the cost of matching the outside offer becomes too expensive even if the agent exerted high effort, and developed human capital. The principal's payoffs from high and low effort, conditional on success, are identical:

$$U_L^L|S = U_L^H|S = \pi_S - c. \tag{3.45}$$

A general ranking is now more difficult to give. We have that $U_L^L|F > U_L^L|S$ as long as $\alpha < \frac{\bar{\psi}}{\Delta p} + c - \Delta\pi$. However, as long as $\alpha < \alpha_3(\psi)$, the previous inequality is satisfied, and $U_L^L|F > U_L^L|S$ holds. In Region 3, the principal's decision rule can be re-written as follows:

$$U_L^L - U_L^H = \Delta p(\pi_F - (\pi_S - c)) - ((1 - p_L)(\alpha - \frac{\bar{\psi}}{\Delta p}) + (1 - p_H)v). \tag{3.46}$$

From (3.46) we can see that the principal's benefit of inducing low effort, relative to high effort, does not depend on α. Her cost of inducing low effort is decreased by $p_H v$, since if the agent exerts effort and succeeds, the principal chooses not to match the outside offer, and does not obtain the human capital benefit. Her cost of inducing low effort is once again decreasing in α at rate $1 - p_L$. Once $\alpha > \alpha_3(\psi)$, inducing low effort is too expensive relative to letting the agent exert effort.

3.4 Conclusion

We analyze a contracting game between one principal and one agent, when the principal may hire the agent to exert high or low effort on a project. If the project succeeds, the agent gets an outside offer, which the principal must match to retain the agent. If the agent exerts high effort, he develops valuable human capital, regardless of the project's outcome.

The agent is self motivated for outside offers large enough, which implies that the principal will have to reward failure in order to get the agent to shirk. We show that when the difference between the probability of success given high and low effort is large enough, the agent can have a high enough cost of effort that the principal may find it profitable to induce low effort, instead of letting him exert low effort like he prefers.

When the average probability of success on the project is greater than one-half, this behavior can create an interesting non-monotonicity in the principal's profit maximizing cost of effort. The principal is best off when the agent fails, so as long as incentivizing low effort is not too expensive, she will do so.

APPENDICES

Appendix A

High cost type's first period incentive constraint

Given the expression for the low and high cost firm's equilibrium efforts, one can verify that the high cost firm's incentive constraint is satisfied in sufficiently noisy environments. Since the high cost type's participation constraint binds in expectation, it is sufficient to check that

$$\underline{t}_1 - \frac{\gamma}{2}\left(\bar{\beta} - \underline{c}_1\right)^2 \leq 0. \tag{A.1}$$

Substituting for $\underline{t}_1$ from (1.28) and simplifying, this requires

$$\frac{\delta}{\gamma \Delta\beta} \int_{\mathbb{R}} \underline{U}_2(\rho_2)(\bar{g} - \underline{g}) dc_1 \leq \bar{c}_1 - \underline{c}_1. \tag{A.2}$$

Now, from (1.33) and (1.34),[1]

$$\begin{aligned} \bar{c}_1 - \underline{c}_1 = & \frac{1+\lambda-\rho}{(1-\rho)(1+\lambda)}\Delta\beta \\ & + \frac{\delta}{\gamma\rho(1-\rho)(1+\lambda)} \frac{d}{d\underline{c}_1}\left[\rho\lambda \int_{\mathbb{R}} \underline{U}_2(\rho_2)(\bar{g} - \underline{g}) dc_1 - E[W_2]\right]. \end{aligned} \tag{A.3}$$

[1] And using the fact that

$$\frac{d}{d\bar{c}_1}\left[\rho\lambda \int_{\mathbb{R}} \underline{U}_2(\rho_2)(\bar{g} - \underline{g}) dc_1 - E[W_2]\right] = -\frac{d}{d\underline{c}_1}\left[\rho\lambda \int_{\mathbb{R}} \underline{U}_2(\rho_2)(\bar{g} - \underline{g}) dc_1 - E[W_2]\right]$$

Thus, the high cost firm's incentive constraint is satisfied when

$$\begin{aligned}\frac{1+\lambda-\rho}{(1-\rho)(1+\lambda)}\Delta\beta \geq & \frac{\delta}{\gamma\rho(1-\rho)(1+\lambda)}\frac{d}{d\underline{c}_1}E[W_2] \\ & -\frac{\delta\lambda}{\gamma(1-\rho)(1+\lambda)}\frac{d}{d\underline{c}_1}\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1 \\ & +\delta\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1. \end{aligned} \tag{A.4}$$

From Proposition 4, $\frac{d}{d\underline{c}_1}E[W_2] < 0$. Therefore, it must be checked that when the variance is sufficiently large,

$$-\frac{\delta\lambda}{\gamma(1-\rho)(1+\lambda)}\frac{d}{d\underline{c}_1}\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1 + \delta\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1 \approx 0. \tag{A.5}$$

From Proposition 3,

$$\frac{d}{d\underline{c}_1}\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1 = \int_{c_1^0}^{\infty}\frac{d\underline{u}_2}{d\rho_2}k\left[\underline{g}'\bar{g}^2-\underline{g}^2\bar{g}'\right]dc_1 + \int_{-\infty}^{c_1^0}\frac{d\underline{u}_2^0}{d\rho_2}k\left[\underline{g}'\bar{g}^2-\underline{g}^2\bar{g}'\right]dc_1. \tag{A.6}$$

As the variance of first period cost increases, the slope of the density goes to zero. As the slope of the density goes to zero, so too does $\frac{d}{d\underline{c}_1}\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1$.

Turning attention to $\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1$, integration by parts yields

$$\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1 = -\left[\int_{-\infty}^{c_1^0}\frac{d\underline{u}_2^0}{d\rho_2}\frac{d\rho_2}{dc_1}\left[\bar{G}-\underline{G}\right]dc_1 + \int_{c_1^0}^{\infty}\frac{d\underline{u}_2}{d\rho_2}\frac{d\rho_2}{dc_1}\left[\bar{G}-\underline{G}\right]dc_1\right]. \tag{A.7}$$

Since $\frac{d\rho_2}{dc_1} = \frac{\rho(1-\rho)[\underline{g}'\bar{g}-\underline{g}\bar{g}']}{D^2}$ goes to zero as the slope of the density goes to zero, this term is close to zero when the variance is large. Thus, the high cost type's incentive constraint is satisfied in noisy enough environments.

Proof of Proposition 1.3

Proof. Consider the expression for the low cost type's first period effort given by (1.33). Abstracting from the effect of the first period contract on expected second period welfare, the low cost type's equilibrium first period effort is less than in a deterministic separating equilibrium (that is, less than $\frac{1}{\gamma}$, the first best) when (1.39) is true. To show that (1.39) holds, consider

$$\frac{d}{d\underline{c}_1}\int_{\mathbb{R}} \underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1 = \frac{d}{d\underline{c}_1}\left[\int_{-\infty}^{c_1^0} \underline{u}_2^0(\bar{g}-\underline{g})dc_1 + \int_{c_1^0}^{\infty} \underline{u}_2(\bar{g}-\underline{g})dc_1\right]$$
$$= \int_{-\infty}^{c_1^0} \frac{d\underline{u}_2^0}{d\rho_2}\frac{d\rho_2}{d\underline{c}_1}(\bar{g}-\underline{g}) + \underline{u}_2^0\underline{g}'dc_1 + \int_{c_1^0}^{\infty} \frac{d\underline{u}_2}{d\rho_2}\frac{d\rho_2}{d\underline{c}_1}(\bar{g}-\underline{g}) + \underline{u}_2\underline{g}'dc_1. \quad \text{(A.8)}$$

Integrate the second term under each integral on the right hand side of (A.8) by parts. Doing so yields

$$\int_{-\infty}^{c_1^0} \frac{d\underline{u}_2^0}{d\rho_2}\left[\frac{d\rho_2}{d\underline{c}_1}\bar{g} - \left(\frac{d\rho_2}{d\underline{c}_1}+\frac{d\rho_2}{dc_1}\right)\underline{g}\right]dc_1 + \underline{u}_2^0\underline{g}\Big|_{-\infty}^{c_1^0}$$
$$+ \int_{c_1^0}^{\infty} \frac{d\underline{u}_2}{d\rho_2}\left[\frac{d\rho_2}{d\underline{c}_1}\bar{g} - \left(\frac{d\rho_2}{d\underline{c}_1}+\frac{d\rho_2}{dc_1}\right)\underline{g}\right]dc_1 + \underline{u}_2\underline{g}\Big|_{c_1^0}^{\infty}. \quad \text{(A.9)}$$

Now,

$$\frac{d\rho_2}{d\underline{c}_1} = \frac{-\rho(1-\rho)\underline{g}'\bar{g}}{D^2}, \quad \text{(A.10)}$$

and

$$\frac{d\rho_2}{dc_1} = \frac{\rho(1-\rho)[\underline{g}'\bar{g} - \underline{g}\bar{g}']}{D^2}, \quad \text{(A.11)}$$

where $D = \rho\underline{g} + (1-\rho)\bar{g}$. Thus,

$$\frac{d\rho_2}{d\underline{c}_1} + \frac{d\rho_2}{dc_1} = \frac{-\rho(1-\rho)\underline{g}\bar{g}'}{D^2}. \tag{A.12}$$

Further,

$$\underline{u}_2^0\underline{g}\Big|_{-\infty}^{c_1^0} + \underline{u}_2\underline{g}\Big|_{c_1^0}^{\infty} = \underline{g}(c_1^0)\left[\underline{u}_2^0(\rho_2^0) - \underline{u}_2(\rho_2^0)\right]. \tag{A.13}$$

When $\rho_2 = \rho_2^0$, it is easily verified that

$$\underline{u}_2^0(\rho_2^0) = \frac{\gamma}{2}\Delta\beta^2 = \underline{u}_2(\rho_2^0). \tag{A.14}$$

After substituting for the relevant terms and simplifying, (A.9) becomes

$$\int_{-\infty}^{c_1^0} \frac{d\underline{u}_2^0}{d\rho_2}k\left[\underline{g}'\bar{g}^2 - \underline{g}^2\bar{g}'\right]dc_1 + \int_{c_1^0}^{\infty} \frac{d\underline{u}_2}{d\rho_2}k\left[\underline{g}'\bar{g}^2 - \underline{g}^2\bar{g}'\right]dc_1, \tag{A.15}$$

where $k = \frac{-\rho(1-\rho)}{D^2}$.

Because $\frac{d\underline{u}_2}{d\rho_2} < 0$ and $\frac{d\underline{u}_2^0}{d\rho_2} < 0$, to show that

$$\underline{g}'\bar{g}^2 - \underline{g}^2\bar{g}' < 0, \quad \forall\, c_1, \tag{A.16}$$

it is sufficient to show that the above integrals are negative over their respective limits of integration. This follows from the monotone likelihood ratio property (see, e.g., the proof of

Theorem 2 in Jeitschko and Mirman (2002)). Thus,

$$\frac{d}{d\underline{c}_1}\rho\lambda\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1$$
$$=\rho\lambda\left[\int_{-\infty}^{c_1^0}\frac{d\underline{u}_2^0}{d\rho_2}k\left[\underline{g}'\bar{g}^2-\underline{g}^2\bar{g}'\right]dc_1+\int_{c_1^0}^{\infty}\frac{d\underline{u}_2}{d\rho_2}k\left[\underline{g}'\bar{g}^2-\underline{g}^2\bar{g}'\right]dc_1\right]<0, \quad \text{(A.17)}$$

and the low cost firm's first period effort is decreased. A similar proof shows that

$$\frac{d}{d\bar{c}_1}\left[\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1\right]=-\frac{d}{d\underline{c}_1}\left[\int_{\mathbb{R}}\underline{U}_2(\rho_2)(\bar{g}-\underline{g})dc_1\right]>0. \quad \text{(A.18)}$$

Thus, the effect of the dynamic portion of the low cost firm's first period transfer is to decrease the distance between cost targets, and reduce how much the regulator updates her prior for any given cost realization. □

Proof of Lemma 1.1

Proof. From the perspective of the second period, expected second period welfare is given by (1.22). When $c_1 > c_1^0$, welfare can be expressed as

$$w_2=\underset{\underline{e}_2,\bar{e}_2}{\operatorname{argmax}}\ S-\rho_2\left((1+\lambda)\left(\underline{\beta}-\underline{e}_2+\frac{\gamma}{2}(\underline{e}_2)^2\right)+\lambda\underline{u}_2\right)$$
$$-(1-\rho_2)(1+\lambda)\big(\bar{\beta}-\bar{e}_2+\frac{\gamma}{2}(\bar{e}_2)^2\big). \quad \text{(A.19)}$$

By the envelope theorem,

$$\frac{dw_2}{d\rho_2} = -(1+\lambda)\big(\underline{\beta} - \underline{e}_2(\rho_2) + \frac{\gamma}{2}(\underline{e}_2(\rho_2))^2\big) - \lambda\underline{u}_2(\bar{e}_2(\rho_2)) + (1+\lambda)\big(\bar{\beta} - \bar{e}_2(\rho_2) + \frac{\gamma}{2}(\bar{e}_2(\rho_2))^2\big)$$
$$= (1+\lambda)\big(\Delta\beta + \frac{1}{2\gamma}\big) - \lambda\underline{u}_2(\bar{e}_2(\rho_2)) - (1+\lambda)\big(\bar{e}_2(\rho_2) - \frac{\gamma}{2}(\bar{e}_2(\rho_2))^2\big). \quad \text{(A.20)}$$

Thus,

$$\frac{d^2w_2}{d\rho_2^2} = -\lambda\frac{d\underline{u}_2}{d\bar{e}_2}\frac{d\bar{e}_2}{d\rho_2} - (1+\lambda)(1-\gamma\bar{e}_2(\rho_2))\frac{d\bar{e}_2}{d\rho_2} > 0, \quad \text{(A.21)}$$

since $\frac{d\underline{u}_2}{d\bar{e}_2} > 0$ and $\frac{d\bar{e}_2}{d\rho_2} < 0$, and the high cost type's effort is less than the first best, which implies $(1-\gamma\bar{e}_2(\rho_2)) > 0$. Because $(1-\gamma\bar{e}_2^0) > 0$ and $\frac{d\underline{u}_2^0}{d\bar{e}_2^0} > 0$ and $\frac{d\bar{e}_2^0}{d\rho_2} < 0$ as well, the proof is identical for w_2^0. Thus, information is valuable. □

Proof of Proposition 1.4

Proof. From the perspective of the first period,

$$E[W_2(\rho_2)] = \int_{-\infty}^{c_1^0} w_2^0\left[\rho\underline{g} + (1-\rho)\bar{g}\right]dc_1 + \int_{c_1^0}^{\infty} w_2\left[\rho\underline{g} + (1-\rho)\bar{g}\right]dc_1. \quad \text{(A.22)}$$

First, consider

$$\frac{dE[W_2(\rho_2)]}{d\underline{c}_1} = \int_{-\infty}^{c_1^0}\frac{dw_2^0}{d\rho_2}\frac{d\rho_2}{d\underline{c}_1}\left[\rho\underline{g} + (1-\rho)\bar{g}\right] - w_2^0\rho\underline{g}'dc_1$$
$$+ \int_{c_1^0}^{\infty}\frac{dw_2}{d\rho_2}\frac{d\rho_2}{d\underline{c}_1}\left[\rho\underline{g} + (1-\rho)\bar{g}\right] - w_2\rho\underline{g}'dc_1. \quad \text{(A.23)}$$

Integrate the second term under each integral by parts. Doing so yields

$$\int_{-\infty}^{c_1^0} \frac{dw_2^0}{d\rho_2}\left[\left(\frac{d\rho_2}{d\underline{c}_1}+\frac{d\rho_2}{dc_1}\right)\rho\underline{g}+\frac{d\rho_2}{d\underline{c}_1}(1-\rho)\bar{g}\right]dc_1 - w_2^0\rho\underline{g}\Big|_{-\infty}^{c_1^0} + \int_{c_1^0}^{\infty} \frac{dw_2}{d\rho_2}\left[\left(\frac{d\rho_2}{d\underline{c}_1}+\frac{d\rho_2}{dc_1}\right)\rho\underline{g}+\frac{d\rho_2}{d\underline{c}_1}(1-\rho)\bar{g}\right]dc_1 - w_2\rho\underline{g}\Big|_{c_1^0}^{\infty}. \quad (A.24)$$

Now,

$$-w_2^0\rho\underline{g}\Big|_{-\infty}^{c_1^0} - w_2\rho\underline{g}\Big|_{c_1^0}^{\infty} = -w_2^0\rho\underline{g}\Big|_{c_1^0} + w_2\rho\underline{g}\Big|_{c_1^0} = 0. \quad (A.25)$$

From the proof of Proposition 3, $\frac{d\rho_2}{d\underline{c}_1} = \frac{-\rho(1-\rho)\underline{g}'\bar{g}}{D^2}$, $\frac{d\rho_2}{dc_1} = \frac{\rho(1-\rho)[\underline{g}'\bar{g}-\underline{g}\bar{g}']}{D^2}$, and $\frac{d\rho_2}{d\underline{c}_1}+\frac{d\rho_2}{dc_1} = \frac{-\rho(1-\rho)\underline{g}\bar{g}'}{D^2}$.

Substituting the above into (A.24) yields

$$\int_{-\infty}^{c_1^0} \frac{dw_2^0}{d\rho_2}\left[\frac{-\rho(1-\rho)\underline{g}\bar{g}'}{D^2}\rho\underline{g} - \frac{\rho(1-\rho)\underline{g}'\bar{g}}{D^2}(1-\rho)\bar{g}\right]dc_1 + \int_{c_1^0}^{\infty} \frac{dw_2}{d\rho_2}\left[\frac{-\rho(1-\rho)\underline{g}\bar{g}'}{D^2}\rho\underline{g} - \frac{\rho(1-\rho)\underline{g}'\bar{g}}{D^2}(1-\rho)\bar{g}\right]dc_1 \quad (A.26)$$

$$= -\left[\int_{-\infty}^{c_1^0} \frac{dw_2^0}{d\rho_2}\rho_2^2(1-\rho)\bar{g}'dc_1 + \int_{-\infty}^{c_1^0} \frac{dw_2^0}{d\rho_2}(1-\rho_2)^2\rho\underline{g}'dc_1\right] - \left[\int_{c_1^0}^{\infty} \frac{dw_2}{d\rho_2}\rho_2^2(1-\rho)\bar{g}'dc_1 + \int_{c_1^0}^{\infty} \frac{dw_2}{d\rho_2}(1-\rho_2)^2\rho\underline{g}'dc_1\right]. \quad (A.27)$$

Using the fact that $(1-\rho_2)^2 = 1-\rho_2-\rho_2(1-\rho_2)$, re-write (A.27) as

$$-\int_{-\infty}^{c_1^0} \frac{dw_2^0}{d\rho_2}\rho_2\left[\rho_2(1-\rho)\bar{g}'-\rho(1-\rho_2)\underline{g}'\right]dc_1 - \int_{-\infty}^{c_1^0}\frac{dw_2^0}{d\rho_2}(1-\rho_2)\rho\underline{g}'dc_1$$
$$-\int_{c_1^0}^{\infty}\frac{dw_2}{d\rho_2}\rho_2\left[\rho_2(1-\rho)\bar{g}'-\rho(1-\rho_2)\underline{g}'\right]dc_1 - \int_{c_1^0}^{\infty}\frac{dw_2}{d\rho_2}(1-\rho_2)\rho\underline{g}'dc_1. \quad \text{(A.28)}$$

Since $\rho_2 = \frac{\rho\underline{g}}{D}$ and $1-\rho_2 = \frac{(1-\rho)\bar{g}}{D}$,

$$\rho_2(1-\rho)\bar{g}-\rho(1-\rho_2)\underline{g}' = \frac{\rho(1-\rho)}{D}[\bar{g}'\underline{g}-\bar{g}\underline{g}'] = -\frac{d\rho_2}{dc_1}D. \quad \text{(A.29)}$$

Thus, (A.28) becomes

$$\int_{-\infty}^{c_1^0}\frac{dw_2^0}{d\rho_2}\frac{d\rho_2}{dc_1}\rho_2 D dc_1 - \int_{-\infty}^{c_1^0}\frac{dw_2^0}{d\rho_2}(1-\rho_2)\rho\underline{g}'dc_1$$
$$+\int_{c_1^0}^{\infty}\frac{dw_2}{d\rho_2}\frac{d\rho_2}{dc_1}\rho_2 D dc_1 - \int_{c_1^0}^{\infty}\frac{dw_2}{d\rho_2}(1-\rho_2)\rho\underline{g}'dc_1. \quad \text{(A.30)}$$

Once again, use the fact that $D\rho_2 = \rho\underline{g}$, and (A.30) becomes

$$\int_{-\infty}^{c_1^0}\frac{dw_2^0}{d\rho_2}\frac{d\rho_2}{dc_1}\rho\underline{g}dc_1 - \int_{-\infty}^{c_1^0}\frac{dw_2^0}{d\rho_2}(1-\rho_2)\rho\underline{g}'dc_1$$
$$+\int_{c_1^0}^{\infty}\frac{dw_2}{d\rho_2}\frac{d\rho_2}{dc_1}\rho\underline{g}dc_1 - \int_{c_1^0}^{\infty}\frac{dw_2}{d\rho_2}(1-\rho_2)\rho\underline{g}'dc_1. \quad \text{(A.31)}$$

Integrate the second and fourth integrals in (A.31) by parts:

$$\int_{-\infty}^{c_1^0}\frac{dw_2^0}{d\rho_2}(1-\rho_2)\rho\underline{g}'dc_1$$
$$= \frac{dw_2^0}{d\rho_2}(1-\rho_2)\rho\underline{g}\Bigg|_{-\infty}^{c_1^0} - \int_{-\infty}^{c_1^0}\left(\frac{d^2w_2^0}{d\rho_2^2}\frac{d\rho_2}{dc_1}(1-\rho_2)-\frac{dw_2^0}{d\rho_2}\frac{d\rho_2}{dc_1}\right)\rho\underline{g}dc_1, \quad \text{(A.32)}$$

and

$$\int_{c_1^0}^{\infty} \frac{dw_2}{d\rho_2}(1-\rho_2)\rho \underline{g}' dc_1$$
$$= \frac{dw_2}{d\rho_2}(1-\rho_2)\rho\underline{g}\bigg|_{c_1^0}^{\infty} - \int_{c_1^0}^{\infty} \left(\frac{d^2 w_2}{d\rho_2^2}\frac{d\rho_2}{dc_1}(1-\rho_2) - \frac{dw_2}{d\rho_2}\frac{d\rho_2}{dc_1} \right) \rho\underline{g} dc_1. \quad \text{(A.33)}$$

Substituting back in to (A.31) yields

$$\frac{dE[W_2(\rho_2)]}{d\underline{c}_1} = \int_{-\infty}^{c_1^0} \frac{d^2 w_2^0}{d\rho_2^2}(1-\rho_2)\frac{d\rho_2}{dc_1}\rho\underline{g} dc_1 + \int_{c_1^0}^{\infty} \frac{d^2 w_2}{d\rho_2^2}(1-\rho_2)\frac{d\rho_2}{dc_1}\rho\underline{g} dc_1$$
$$+ (1-\rho_2)\rho\underline{g}\Big(\frac{dw_2}{d\rho_2} - \frac{dw_2^0}{d\rho_2}\Big)\bigg|_{c_1^0}. \quad \text{(A.34)}$$

Since $\frac{d\rho_2}{dc_1} < 0$ by the monotone likelihood ratio property, by Lemma 1 the integrals are negative for all c_1. It is left to show that, when evaluated at c_1^0,

$$\frac{dw_2}{d\rho_2} - \frac{dw_2^0}{d\rho_2} = 0. \quad \text{(A.35)}$$

Lemma 1 gives the expression for $\frac{dw_2}{d\rho_2}$, and a similar argument yields

$$\frac{dw_2^0}{d\rho_2} = (1+\lambda)\big(\Delta\beta + \frac{1}{2\gamma}\big) - \lambda \underline{u}_2^0(\rho_2) - (1+\lambda)\big(\bar{e}_2^0(\rho_2) - \frac{\gamma}{2}(\bar{e}_2^0(\rho_2))^2\big). \quad \text{(A.36)}$$

Thus, when evaluated at c_1^0,

$$\frac{dw_2}{d\rho_2} - \frac{dw_2^0}{d\rho_2} = \lambda \left[\underline{u}_2^0(\rho_2^0) - \underline{u}_2(\rho_2^0) \right]$$
$$+ (1+\lambda)\left[\bar{e}_2^0(\rho_2^0) - \bar{e}_2(\rho_2^0) + \frac{\gamma}{2}(\bar{e}_2(\rho_2^0))^2 - \frac{\gamma}{2}(\bar{e}_2^0(\rho_2^0))^2 \right]. \quad \text{(A.37)}$$

From Proposition 1, $\underline{u}_2^0(\rho_2^0) - \underline{u}_2(\rho_2^0) = 0$. Further,

$$\bar{e}_2^0(\rho_2^0) = \Delta\beta = \bar{e}_2(\rho_2^0). \tag{A.38}$$

Thus,

$$\frac{dw_2}{d\rho_2} - \frac{dw_2^0}{d\rho_2} = 0, \tag{A.39}$$

and

$$\frac{dE[W_2(\rho_2)]}{d\underline{c}_1} = \int_{-\infty}^{c_1^0} \frac{d^2 w_2^0}{d\rho_2^2}(1-\rho_2)\frac{d\rho_2}{dc_1}\rho\underline{g}dc_1 + \int_{c_1^0}^{\infty} \frac{d^2 w_2}{d\rho_2^2}(1-\rho_2)\frac{d\rho_2}{dc_1}\rho\underline{g}dc_1 < 0. \tag{A.40}$$

A similar proof shows that

$$\frac{dE[W_2(\rho_2)]}{d\bar{c}_1} = -\left[\int_{-\infty}^{c_1^0} \frac{d^2 w_2^0}{d\rho_2^2}(1-\rho_2)\frac{d\rho_2}{dc_1}\rho\underline{g}dc_1 + \int_{c_1^0}^{\infty} \frac{d^2 w_2}{d\rho_2^2}(1-\rho_2)\frac{d\rho_2}{dc_1}\rho\underline{g}dc_1\right] > 0. \tag{A.41}$$

Thus, the effect of expected second period welfare is to increase the distance between the first period cost targets. □

Proof of Proposition 1.5

Proof. To prove Proposition 1.5, consider

$$\frac{d}{d\underline{c}_1}\left[\int_{-\infty}^{c_1^0} A\,\bar{g}\,dc_1 + \int_{c_1^0}^{\infty} B\,\bar{g}\,dc_1\right] = \int_{-\infty}^{c_1^0} A'\frac{d\bar{e}_2^0}{d\rho_2}\frac{d\rho_2}{d\underline{c}_1}\bar{g}dc_1 + \int_{c_1^0}^{\infty} B'\frac{d\bar{e}_2}{d\rho_2}\frac{d\rho_2}{d\underline{c}_1}\bar{g}dc_1, \tag{A.42}$$

where

$$A' = (1-\rho)(1+\lambda) - (1+\lambda-\rho)\gamma\bar{e}_2^0, \tag{A.43}$$

and

$$B' = (1-\rho)(1+\lambda)(1-\gamma\bar{e}_2) - \rho\lambda\gamma\Delta\beta. \tag{A.44}$$

First, focus on

$$\int_{-\infty}^{c_1^0} A' \frac{d\bar{e}_2^0}{d\rho_2} \frac{d\rho_2}{d\underline{c}_1} \bar{g} dc_1. \tag{A.45}$$

Since

$$\gamma\bar{e}_2^0 = \frac{(1-\rho_2)(1+\lambda)}{1+\lambda-\rho_2}, \tag{A.46}$$

it follows that

$$A' = \frac{1+\lambda}{1+\lambda-\rho_2}\big[(1-\rho)(1+\lambda-\rho_2) - (1-\rho_2)(1+\lambda-\rho)\big] = \frac{\lambda(1+\lambda)(\rho_2-\rho)}{1+\lambda-\rho_2}. \tag{A.47}$$

Since $c_1 \leq c_1^0 < \hat{c}_1$ (see Figure 1), $\rho_2 > \rho$. Thus , $A' > 0$.

Further, for every $c_1 \leq \underline{c}_1$, $\underline{g}$ is increasing, so $\underline{g}' \geq 0$ (see Figure 1); thus,

$$\frac{d\rho_2}{d\underline{c}_1} = \frac{-\rho(1-\rho)\underline{g}'\bar{g}}{D^2} < 0. \tag{A.48}$$

Therefore, since $\frac{d\bar{e}_2^0}{d\rho_2} < 0$ for all c_1, and since $c_1^0 \leq \underline{c}_1$,

$$\int_{-\infty}^{c_1^0} A' \frac{d\bar{e}_2^0}{d\rho_2} \frac{d\rho_2}{d\underline{c}_1} \bar{g} dc_1 > 0 \tag{A.49}$$

for $c_1 \in \big(-\infty, c_1^0\big]$.

Now, return attention to

$$\int_{c_1^0}^{\infty} B' \frac{d\bar{e}_2}{d\rho_2} \frac{d\rho_2}{d\underline{c}_1} \bar{g} dc_1. \tag{A.50}$$

Using the definition of $\frac{d\rho_2}{d\underline{c}_1}$, (A.50) can be re-written

$$\frac{-\rho}{1-\rho}\int_{c_1^0}^{\infty} B'\frac{d\bar{e}_2}{d\rho_2}(1-\rho_2)^2\underline{g}'dc_1. \tag{A.51}$$

Integrating by parts yields

$$B'\frac{d\bar{e}_2}{d\rho_2}(1-\rho_2)^2\underline{g}\Big|_{c_1^0}^{\infty}$$
$$-\int_{c_1^0}^{\infty}\left[-(1-\rho)(1+\lambda)\gamma\left(\frac{d\bar{e}_2}{d\rho_2}\right)^2(1-\rho_2)+B'\left[\frac{d^2\bar{e}_2}{d\rho_2^2}(1-\rho_2)-2\frac{d\bar{e}_2}{d\rho_2}\right]\right](1-\rho_2)\frac{d\rho_2}{dc_1}\underline{g}dc_1. \tag{A.52}$$

Notice that

$$\frac{d^2\bar{e}_2}{d\rho_2^2}(1-\rho_2)-2\frac{d\bar{e}_2}{d\rho_2}=\frac{-2\lambda\Delta\beta(1-\rho_2)}{(1-\rho_2)^3(1+\lambda)}-\frac{-2\lambda\Delta\beta}{(1-\rho_2)^2(1+\lambda)}=0. \tag{A.53}$$

Thus, (A.51) becomes

$$\frac{-\rho}{1-\rho}\left[-B'\frac{d\bar{e}_2}{d\rho_2}(1-\rho_2)^2\underline{g}\Big|_{c_1^0}+\int_{c_1^0}^{\infty}(1-\rho)(1+\lambda)\gamma\left(\frac{d\bar{e}_2}{d\rho_2}\right)^2(1-\rho_2)^2\frac{d\rho_2}{dc_1}\underline{g}dc_1\right]. \tag{A.54}$$

First, notice that

$$\frac{-\rho}{1-\rho}\int_{c_1^0}^{\infty}(1-\rho)(1+\lambda)\gamma\left(\frac{d\bar{e}_2}{d\rho_2}\right)^2(1-\rho_2)^2\frac{d\rho_2}{dc_1}\underline{g}dc_1>0, \tag{A.55}$$

since $\frac{d\rho_2}{dc_1}<0$ for all c_1, and every other term under the integral in (A.55) is positive. Now, consider

$$\frac{\rho}{1-\rho}\, B' \frac{d\bar{e}_2}{d\rho_2}(1-\rho_2)^2 \underline{g}\Big|_{c_1^0}. \tag{A.56}$$

Since

$$\frac{d\bar{e}_2}{d\rho_2} = \frac{-\lambda\Delta\beta}{(1-\rho_2)^2(1+\lambda)}, \tag{A.57}$$

(A.56) can be simplified to

$$\frac{-\rho\lambda\Delta\beta}{(1-\rho)(1+\lambda)} B'(c_1^0) g(c_1^0). \tag{A.58}$$

When evaluated at $c_1 = c_1^0$, $\bar{e}_2 = \Delta\beta$. Thus,

$$B'(c_1^0) = (1-\rho)(1+\lambda)\left[1 - \frac{1+\lambda-\rho}{(1-\rho)(1+\lambda)}\gamma\Delta\beta\right], \tag{A.59}$$

and (A.56) further simplifies to

$$-\rho\lambda\Delta\beta\left[1 - \frac{1+\lambda-\rho}{(1-\rho)(1+\lambda)}\gamma\Delta\beta\right]\underline{g}(c_1^0). \tag{A.60}$$

Clearly, the term in brackets in (A.60) is less than one. It is also equal to $\gamma(\bar{e}_2(\rho) - \Delta\beta)$, where $\bar{e}_2(\rho) - \Delta\beta$ is the low cost firm's effort from mimicking the high cost firm in a static game in which the regulator's beliefs are given by ρ. This is assumed to be positive; thus, the expression given in (A.60) is negative. However, the terms multiplying $\underline{g}(c_1^0)$ are small, and if $\underline{g}(c_1^0) \approx 0$, the term in (A.60) can be ignored in signing the first order condition.

Thus,

$$\frac{d}{d\underline{c}_1}\left[\int_{-\infty}^{c_1^0} A\,\bar{g}\,dc_1 + \int_{c_1^0}^{\infty} B\,\bar{g}\,dc_1\right] \approx \int_{-\infty}^{c_1^0} A' \frac{d\bar{e}_2^0}{d\rho_2}\frac{d\rho_2}{d\underline{c}_1}\bar{g}dc_1$$

$$- \frac{\rho}{1-\rho}\int_{c_1^0}^{\infty}(1-\rho)(1+\lambda)\gamma\left(\frac{d\bar{e}_2}{d\rho_2}\right)^2(1-\rho_2)^2\frac{d\rho_2}{dc_1}\underline{g}dc_1 > 0, \quad \text{(A.61)}$$

and the desired result is obtained. A similar proof shows that

$$\frac{d}{d\bar{c}_1}\left[\int_{-\infty}^{c_1^0} A\,\bar{g}\,dc_1 + \int_{c_1^0}^{\infty} B\,\bar{g}\,dc_1\right] = -\frac{d}{d\underline{c}_1}\left[\int_{-\infty}^{c_1^0} A\,\bar{g}\,dc_1 + \int_{c_1^0}^{\infty} B\,\bar{g}\,dc_1\right]. \quad \text{(A.62)}$$

Thus, the low cost (high cost) firm's effort in the first period is below (above) the commitment optimum, and his effort is increased (decreased) over the course of the interaction with the regulator. □

Appendix B

Existence of a reward schedule that implements output targets in each period

There exists $\hat{y}_1$ such that the following is an equilibrium reward schedule:

$$r_1(y_1) = \begin{cases} \left(\frac{y_1}{\underline{\theta}}\right)^2 - \delta \underline{E}_1[\overline{u}_2(\rho_2) |\, \underline{y}_1], & \text{if} \quad y_1 \in [\underline{y}_1 - \eta,\, \hat{y}_1) \\ \left(\frac{y_1}{\underline{\theta}}\right)^2 - \delta \underline{E}_1[\overline{u}_2(\rho_2) |\, \underline{y}_1] + & \\ \frac{1}{P(y_1 \geq \hat{y}_1)} \left[\frac{\overline{y}_1^2 - \underline{y}_1^2}{\overline{\theta}^2} + \delta\alpha \left(\frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right) (\overline{u}_2(\rho_l) - \overline{u}_2(\rho_h))\right], & \text{if} \quad y_1 \in [\hat{y}_1,\, \overline{y}_1 + \eta] \end{cases}$$

To see this, note that the principal punishes output realizations less than $\underline{y}_1 - \eta$. Therefore, the low productivity agent does not shirk. To see that the high productivity agent is strictly worse off from targeting $y_1 \notin \{\underline{y}_1, \overline{y}_1\}$, note that due to the uniform noise assumption, the probability that he receives the bonus increases linearly in his choice of output target. Because his disutility from effort is convex, however, the increase in his expected transfer from a slight increase in effort is outweighed by the increase in his disutility from effort.

This reward schedule and the second period reward schedule $r_2(y_2)$ are a straightforward extension of Proposition 4 in [24]. Therefore, see [24] for further discussion.

The low ability agent's first period incentive constraint is slack

Since the low ability agent's expected utility from targeting $\underline{y}_1$ is equal to his outside option of zero, it suffices to show that there exists *eta* large enough such that

$$\overline{r}_1 - \left(\frac{\overline{y}_1}{\underline{\theta}}\right)^2 \leq 0. \tag{B.1}$$

After some algebra, one can show that this is equivalent to

$$\frac{\delta}{2\eta}\left(\frac{\underline{\theta}^2}{2}\right)^2 \left[\alpha\left(C^2(1-\alpha) - C^2(\alpha)\right) - (1-\alpha)\left(C^2(1-\alpha) - C^2(\rho_m)\right)\right] \\ - \frac{1}{\rho}\frac{\overline{\theta}^2}{2}\frac{\delta}{2\eta}A + \frac{1}{1-\rho\Theta^2}\frac{\underline{\theta}^2}{2}\frac{\delta}{2\eta}A \leq \frac{\overline{\theta}^2}{2} + C(\rho)\frac{\underline{\theta}^2}{2}. \tag{B.2}$$

The right hand side does not depend on η, and the left hand side goes to zero as η grows.

Proof of Proposition 2.1

Proof. First, note that

$$\frac{d\,\overline{r}_1^D}{d(\overline{y}_1 - \underline{y}_1)} = \frac{\delta}{2\eta}\big[\alpha(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_h)) - (1-\alpha)(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_m))\big]. \tag{B.3}$$

Sufficient for this term to be positive is

$$\frac{\alpha}{1-\alpha} > \frac{\overline{u}_2(\rho_l) - \overline{u}_2(\rho_m)}{\overline{u}_2(\rho_l) - \overline{u}_2(\rho_h)}. \tag{B.4}$$

By Corollary 1, the right hand side of (B.4) is always less than one, and the left hand side is greater than or equal to one for all $\alpha \in [\frac{1}{2}, 1]$. Therefore, $\frac{d\,\overline{r}_1^D}{d(\overline{y}_1 - \underline{y}_1)} > 0$ as desired. □

Proof of Proposition 2.2

Proof. From (2.41),

$$\frac{d\,\underline{r}_1^D}{d(\overline{y}_1 - \underline{y}_1)} = -\frac{\delta(1-\alpha)}{2\eta}\left(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_m)\right). \tag{B.5}$$

By Corollary 1, $\overline{u}_2(\rho_l) - \overline{u}_2(\rho_m) > 0$, and thus $\frac{d\,\underline{r}_1^D}{d(\overline{y}_1 - \underline{y}_1)} < 0$, as desired. □

Proof of Lemma 2.2

The principal's expected second period payoff can be expressed as follows:

$$v_2 = \underset{\underline{y}_2, \overline{y}_2}{\text{argmax}}\ \rho_2\left[\overline{y}_2 - \left(\frac{\overline{y}_2}{\overline{\theta}}\right)^2 - \left(\frac{\underline{y}_2}{\underline{\theta}}\right)^2 + \left(\frac{\underline{y}_2}{\overline{\theta}}\right)^2\right] + (1-\rho_2)\left[\underline{y}_2 - \left(\frac{\underline{y}_2}{\underline{\theta}}\right)^2\right]. \tag{B.6}$$

By the envelope theorem,

$$\frac{dv_2}{d\rho_2} = \overline{y}_2(\rho_2) - \left(\frac{\overline{y}_2(\rho_2)}{\overline{\theta}}\right)^2 - \left(\frac{\underline{y}_2(\rho_2)}{\underline{\theta}}\right)^2 + \left(\frac{\underline{y}_2(\rho_2)}{\overline{\theta}}\right)^2 - \left[\underline{y}_2(\rho_2) - \left(\frac{\underline{y}_2(\rho_2)}{\underline{\theta}}\right)^2\right]. \tag{B.7}$$

In the second period, the optimal output target for the high productivity firm does not depend on second period beliefs ($\overline{y}_2 = \overline{\theta}^2/2$), and the low productivity worker's optimal output target as a function of second period beliefs is given by (2.14). Thus,

$$\frac{dv_2}{d\rho_2} = \frac{\overline{\theta}^2}{4} - C^2(\rho_2)\frac{\underline{\theta}^2}{4}\left[1 - \Theta^2\right] - C(\rho_2)\frac{\underline{\theta}^2}{2}\left[1 - \frac{C(\rho_2)}{2}\right], \tag{B.8}$$

and

$$\begin{aligned}\frac{d^2v_2}{d\rho_2^2} &= -2C(\rho_2)C'(\rho_2)\frac{\underline{\theta}^2}{4}\left[1 - \Theta^2\right] - \frac{\underline{\theta}^2}{2}\left[C'(\rho_2)\left(1 - \frac{C(\rho_2)}{2}\right) - C(\rho_2)\frac{C'(\rho_2)}{2}\right] \\ &= -2C(\rho_2)C'(\rho_2)\frac{\underline{\theta}^2}{4}\left[1 - \Theta^2\right] - C'(\rho_2)\frac{\underline{\theta}^2}{2}\left[1 - C(\rho_2)\right] > 0, \end{aligned} \tag{B.9}$$

since $C'(\rho_2) < 0$, and $1 - C(\rho_2) > 0$. Thus, the principal's expected second period payoff is convex in second period beliefs.

Proof of Proposition 2.3

Proof. With $E[v_2(\rho_2)]$ given in (2.23),

$$\frac{dE[v_2(\rho_2)]}{d(\overline{y}_1 - \underline{y}_1)} = \frac{1}{2\eta}\left[\rho v_2(\rho_h) + (1-\rho)v_2(\rho_l) - v_2(\rho_m)\right]. \tag{B.10}$$

Sufficient for this to be positive is

$$\rho v_2(\rho_h) + (1-\rho)v_2(\rho_l) \geq v_2(\rho_m)). \tag{B.11}$$

Using the definition of ρ_h, ρ_m and ρ_l,

$$\begin{aligned}\rho v_2(\rho_h) + (1-\rho)v_2(\rho_l) &= \rho v_2(\alpha) + (1-\rho)v_2(1-\alpha) \\ &\geq v_2(\rho\alpha + (1-\rho)(1-\alpha)) \\ &= v_2(\rho_m), \end{aligned} \tag{B.12}$$

where the inequality holds due to Lemma 2.2. Hence the desired result. □

Proof of Proposition 2.4

Proof. First, notice that

$$\begin{aligned}\rho\overline{r}_1 + (1-\rho)\underline{r}_1 &= \rho\left(\overline{r}_1^S + \overline{r}_1^D\right) + (1-\rho)\left(\underline{r}_1^S + \underline{r}_1^D\right) \\ &= \rho\overline{r}_1^S + (1-\rho)\underline{r}_1^S - \delta(1-\alpha)\overline{u}_2(\rho_2) \\ &+ \delta\left(\frac{\overline{y}_1 - \underline{y}_1}{2\eta}\right)\left(\rho\alpha(\overline{u}_2(1-\alpha) - \overline{u}_2(\alpha)) - (1-\alpha)(\overline{u}_2(1-\alpha) - \overline{u}_2(\rho_2^2))\right). \end{aligned} \tag{B.13}$$

Therefore,

$$\frac{dE[r_1]}{d(\overline{y}_1 - \underline{y}_1)} = \rho\alpha(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_h)) - (1-\alpha)(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_m)), \tag{B.14}$$

and sufficient for $\frac{dE[r_1]}{d(\overline{y}_1 - \underline{y}_1)} > 0$ is

$$\rho\alpha\left(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_h)\right) - (1-\alpha)\left(\overline{u}_2(\rho_l) - \overline{u}_2(\rho_m)\right) \geq 0. \tag{B.15}$$

Recall that for generic second period beliefs ρ_2, the high productivity agent receives expected second period rent

$$\overline{u}_2(\rho_2) = C^2(\rho_2)\frac{\underline{\theta}^2}{4}\left[1-\Theta^2\right], \tag{B.16}$$

where $C(\rho_2) = \frac{1-\rho_2}{1-\rho_2\Theta^2}$, and $\Theta = \frac{\underline{\theta}}{\overline{\theta}}$. Thus,

$$\overline{u}_2(\rho_l) - \overline{u}_2(\rho_h) = \frac{\underline{\theta}^2}{4}\left[1-\Theta^2\right]\left(C^2(1-\alpha) - C^2(\alpha)\right), \tag{B.17}$$

and

$$\overline{u}_2(\rho_l) - \overline{u}_2(\rho_m) = \frac{\underline{\theta}^2}{4}\left[1-\Theta^2\right]\left(C^2(1-\alpha) - C^2(\rho_2^2)\right). \tag{B.18}$$

Thus, we wish to show that

$$\rho\alpha\left(C^2(1-\alpha) - C^2(\alpha)\right) > (1-\alpha)\left(C^2(1-\alpha) - C^2(\rho_2^2)\right). \tag{B.19}$$

Note that

$$C^2(1-\alpha) - C^2(\alpha) = \left(C(1-\alpha) - C(\alpha)\right)\cdot\left(C(1-\alpha) + C(\alpha)\right), \tag{B.20}$$

and

$$C^2(1-\alpha) - C^2(\rho_2^2) = \left(C(1-\alpha) - C(\rho_2^2)\right)\cdot\left(C(1-\alpha) + C(\rho_2^2)\right). \tag{B.21}$$

One can show that

$$C(1-\alpha) - C(\alpha) = \frac{(1-\Theta^2)(2\alpha-1)}{(1-(1-\alpha)\Theta^2)(1-\alpha\Theta^2)} \tag{B.22}$$

and

$$C(1-\alpha) - C(\rho_2^2) = \frac{\rho(1-\Theta^2)(2\alpha-1)}{(1-(1-\alpha)\Theta^2)(1-\rho_2^2\Theta^2)}, \tag{B.23}$$

Thus, (B.19) is equivalent to

$$\frac{\alpha}{1-\alpha\Theta^2}\left(C(1-\alpha)+C(\alpha)\right) \geq \frac{1-\alpha}{1-\rho_2^2\Theta^2}\left(C(1-\alpha)+C(\rho_2^2)\right). \tag{B.24}$$

This statement is true if

$$C(1-\alpha)\left[\frac{\alpha}{1-\alpha\Theta^2} - \frac{1-\alpha}{1-\rho_2^2\Theta^2}\right] + (1-\alpha)\left[\frac{\alpha}{(1-\alpha\Theta^2)^2} - \frac{1-\rho_2^2}{(1-\rho_2^2\Theta^2)^2}\right] \geq 0. \tag{B.25}$$

First, notice that

$$\frac{\alpha}{1-\alpha\Theta^2} \geq \frac{1-\alpha}{1-\alpha\Theta^2} \geq \frac{1-\alpha}{1-\rho_2^2\Theta^2}. \tag{B.26}$$

Both inequalities inequality hold since since $\alpha \geq 1/2$ (for the second inequality, note that $\alpha \geq 1/2$ implies that $\alpha \geq \rho_2^2$, which in turn implies that $1-\alpha\Theta^2 \leq 1-\rho_2^2\Theta^2$). Similar logic ($\alpha \geq 1/2 \Rightarrow \alpha \geq 1-\rho_2^2$) shows that

$$\frac{\alpha}{(1-\alpha\Theta^2)^2} \geq \frac{1-\rho_2^2}{(1-\rho_2^2\Theta^2)^2}. \tag{B.27}$$

Thus,

$$\rho\alpha(\overline{u}_2(1-\alpha) - \overline{u}_2(\alpha)) - (1-\alpha)(\overline{u}_2(1-\alpha) - \overline{u}_2(\rho_2^2)) \geq 0 \tag{B.28}$$

for every $\alpha \in [1/2, 1]$, and $\frac{dE[r_1]}{d(\overline{y}_1 - \underline{y}_1)} > 0$ as desired. The principal can decrease the total expected transfer by decreasing the distance between the first period cost targets. $\square$

Proof of Proposition 2.5

Proof. First, consider the high productivity agent's first period output target. Sufficient for $\overline{y}_1 \leq \overline{y}^c$ for all α is to show that $A < 0$ for all α, where A is given in (2.32). The first task is to show that

$$A = B \cdot \Big((1-\rho)(2 - C(\alpha)) - (1 - \rho\alpha\Theta^2)(C(1-\alpha) + C(\rho_2^2)) \Big) , \tag{B.29}$$

where

$$B := \frac{\rho(1-\Theta^2)^2(2\alpha - 1)^2}{(1-\rho_2^2\Theta^2)(1-(1-\alpha)\Theta^2)(1-\alpha\Theta^2)}. \tag{B.30}$$

To see this, note that for generic second period beliefs ρ_2, the principal's second period payoff, given in (2.21), can be re-written

$$v_2(\rho_2) = \rho_2 \frac{\overline{\theta}^2}{4} + (1-\rho_2)C(\rho_2)\frac{\underline{\theta}^2}{4}. \tag{B.31}$$

Thus,

$$\begin{aligned} &\rho v_2(\alpha) + (1-\rho)v_2(1-\alpha) - v_2(\rho_2^2) \\ &\qquad = \frac{\underline{\theta}^2}{4}\Big[\rho(1-\alpha)C(\alpha) + (1-\rho)\alpha C(1-\alpha) - (1-\rho_2^2)C(\rho_2^2)\Big] \\ &\qquad = \frac{\underline{\theta}^2}{4}\Big[\alpha(1-\rho)\Big(C(1-\alpha) - C(\rho_2^2)\Big) - \rho(1-\alpha)\Big(C(\rho_2^2) - C(\alpha)\Big)\Big]. \end{aligned} \tag{B.32}$$

Next, consider

$$\rho\alpha[\overline{u}_2(1-\alpha) - \overline{u}_2(\alpha)] - (1-\alpha)[\overline{u}_2(1-\alpha) - \overline{u}_2(\rho_2^2)]$$
$$= \frac{\theta^2}{4}\left(1-\Theta^2\right)\left[\rho\alpha(C^2(1-\alpha) - C^2(\alpha)) - (1-\alpha)(C^2(1-\alpha) - C^2(\rho_2^2))\right]. \quad \text{(B.33)}$$

Ignoring $\frac{\theta^2}{4}$, which factors out, A becomes

$$\alpha(1-\rho)\left(C(1-\alpha) - C(\rho_2^2)\right) - \rho(1-\alpha)\left(C(\rho_2^2) - C(\alpha)\right)$$
$$-\left(1-\Theta^2\right)\left[\rho\alpha(C^2(1-\alpha) - C^2(\alpha)) - (1-\alpha)(C^2(1-\alpha) - C^2(\rho_2^2))\right]. \quad \text{(B.34)}$$

Consider the term in brackets on the second line of (B.34):

$$\rho\alpha(C^2(1-\alpha) - C^2(\alpha)) - (1-\alpha)(C^2(1-\alpha) - C^2(\rho_2^2))$$
$$= \rho\left(\alpha(C^2(1-\alpha) - C^2(\alpha))) - (1-\alpha)(C^2(1-\alpha) - C^2(\rho_2^2)\right)$$
$$-(1-\rho)(1-\alpha)(C^2(1-\alpha) - C^2(\rho_2^2)$$
$$= \rho\left(\alpha(C^2(1-\alpha) - C^2(\rho_2^2)) + \alpha(C^2(\rho_2^2) - C^2(\alpha)) - (1-\alpha)(C^2(1-\alpha) - C^2(\rho_2^2))\right)$$
$$-(1-\rho)(1-\alpha)(C^2(1-\alpha) - C^2(\rho_2^2)$$
$$= \rho(2\alpha-1)\left(C^2(1-\alpha) - C^2(\rho_2^2)\right) + \rho\alpha\left(C^2(\rho_2^2) - C^2(\alpha)\right)$$
$$-(1-\rho)(1-\alpha)\left(C^2(1-\alpha) - C^2(\rho_2^2)\right). \quad \text{(B.35)}$$

Therefore, (B.34) is equivalent to

$$\alpha(1-\rho)\left(C(1-\alpha)-C(\rho_2^2)\right)-\rho(1-\alpha)\left(C(\rho_2^2)-C(\alpha)\right) \\ -(1-\Theta^2)\rho(2\alpha-1)\left(C^2(1-\alpha)-C^2(\rho_2^2)\right)-(1-\Theta^2)\rho\alpha\left(C^2(\rho_2^2)-C^2(\alpha)\right) \\ +(1-\Theta)^2(1-\rho)(1-\alpha)\left(C^2(1-\alpha)-C^2(\rho_2^2)\right). \quad \text{(B.36)}$$

Next, use the fact that $x^2-y^2=(x-y)(x+y)$ combined with the fact that

$$C(1-\alpha)-C(\rho_2^2)=\frac{\rho(1-\Theta^2)(2\alpha-1)}{(1-(1-\alpha)\Theta^2)(1-\rho_2^2\Theta^2)} \quad \text{(B.37)}$$

and

$$C(\rho_2^2)-C(\alpha)=\frac{(1-\rho)(1-\Theta^2)(2\alpha-1)}{(1-\alpha\Theta^2)(1-\rho_2^2\Theta^2)}, \quad \text{(B.38)}$$

and (B.36) becomes

$$\frac{\rho(1-\rho)(1-\Theta^2)(2\alpha-1)}{(1-\rho_2^2\Theta^2)}\left(C(1-\alpha)-C(\alpha)\right) \\ -\frac{\rho^2(1-\Theta^2)^2(2\alpha-1)^2}{(1-(1-\alpha)\Theta^2)(1-\rho_2^2\Theta^2)}\left(C(1-\alpha)+C(\rho_2^2)\right) \\ -\frac{\rho(1-\rho)(1-\Theta^2)(2\alpha-1)}{(1-\rho_2^2\Theta^2)}\left(\frac{\alpha(1-\Theta^2)}{1-\alpha\Theta^2}(C(\rho_2^2)+C(\alpha))\right) \\ +\frac{\rho(1-\rho)(1-\Theta^2)(2\alpha-1)}{(1-\rho_2^2\Theta^2)}\left(\frac{(1-\alpha)(1-\Theta^2)}{1-(1-\alpha)\Theta^2)}(C(1-\alpha)+C(\rho_2^2))\right). \quad \text{(B.39)}$$

Notice that

$$\frac{\alpha(1-\Theta^2)}{1-\alpha\Theta^2}=1-C(\alpha) \quad \text{(B.40)}$$

and

$$\frac{(1-\alpha)(1-\Theta^2)}{1-(1-\alpha)\Theta^2)} = 1 - C(1-\alpha). \tag{B.41}$$

Therefore, the expression in (B.39) becomes

$$\begin{aligned}
&\frac{\rho(1-\rho)(1-\Theta^2)(2\alpha-1)}{(1-\rho_2^2\Theta^2)}\left(C(1-\alpha)-C(\alpha)\right) \\
&\qquad - \frac{\rho^2(1-\Theta^2)^2(2\alpha-1)^2}{(1-(1-\alpha)\Theta^2)(1-\rho_2^2\Theta^2)}\Big(C(1-\alpha)+C(\rho_2^2)\Big) \\
&+ \frac{\rho(1-\rho)(1-\Theta^2)(2\alpha-1)}{(1-\rho_2^2\Theta^2)}\left(C(1-\alpha)-C(\alpha)\right)\Big(1-C(\alpha)-C(1-\alpha)-C(\rho_2^2)\Big).
\end{aligned} \tag{B.42}$$

Combining terms with like denominators, this is equivalent to

$$\begin{aligned}
&\frac{\rho(1-\rho)(1-\Theta^2)(2\alpha-1)}{(1-\rho_2^2\Theta^2)}\left(C(1-\alpha)-C(\alpha)\right)\Big(2-C(\alpha)-C(1-\alpha)-C(\rho_2^2)\Big) \\
&\qquad - \frac{\rho^2(1-\Theta^2)^2(2\alpha-1)^2}{(1-(1-\alpha)\Theta^2)(1-\rho_2^2\Theta^2)}\Big(C(1-\alpha)+C(\rho_2^2)\Big).
\end{aligned} \tag{B.43}$$

Next, use the fact that

$$C(1-\alpha) - C(\alpha) = \frac{(1-\Theta^2)(2\alpha-1)}{(1-(1-\alpha)\Theta^2)(1-\alpha\Theta^2)}, \tag{B.44}$$

and (B.43) becomes

$$\begin{aligned}
&\frac{\rho(1-\rho)(1-\Theta^2)^2(2\alpha-1)^2}{(1-\rho_2^2\Theta^2)(1-(1-\alpha)\Theta^2)(1-\alpha\Theta^2)}\Big(2-C(\alpha)-C(1-\alpha)-C(\rho_2^2)\Big) \\
&\qquad - \frac{\rho^2(1-\Theta^2)^2(2\alpha-1)^2}{(1-(1-\alpha)\Theta^2)(1-\rho_2^2\Theta^2)}\Big(C(1-\alpha)+C(\rho_2^2)\Big).
\end{aligned} \tag{B.45}$$

Define

$$B := \frac{\rho(1-\Theta^2)^2(2\alpha-1)^2}{(1-\rho_2^2\Theta^2)(1-(1-\alpha)\Theta^2)(1-\alpha\Theta^2)}, \tag{B.46}$$

and (B.45) becomes

$$B \cdot \Big((1-\rho)(2-C(\alpha)-C(1-\alpha)-C(\rho_2^2)) - \rho(1-\alpha\Theta^2)(C(1-\alpha)+C(\rho_2^2))\Big), \tag{B.47}$$

which simplifies to

$$B \cdot \Big((1-\rho)(2-C(\alpha)) - (1-\rho\alpha\Theta^2)(C(1-\alpha)+C(\rho_2^2))\Big). \tag{B.48}$$

Thus,

$$A = B \cdot \Big((1-\rho)(2-C(\alpha)) - (1-\rho\alpha\Theta^2)(C(1-\alpha)+C(\rho_2^2))\Big). \tag{B.49}$$

With this result in hand, notice that $B \geq 0$ for all $\alpha \in [1/2,\ 1]$, $\rho \in (0,1)$, and $\Theta^2 \in (0,1)$. Therefore, it is left to show that if $\Theta^2 \geq 1/2$ or $\rho \geq 1/3$, then

$$(1-\rho)(2-C(\alpha)) - (1-\rho\alpha\Theta^2)(C(1-\alpha)+C(\rho_2^2)) \leq 0 \tag{B.50}$$

for all $\alpha \in [1/2,\ 1]$. Define

$$f(\alpha) := (1-\rho)(2-C(\alpha)) - (1-\rho\alpha\Theta^2)(C(1-\alpha)+C(\rho_2^2)). \tag{B.51}$$

Notice that

$$f(1) = (1-\rho)2 - (1-\rho\Theta^2)\left(1 + \frac{1-\rho}{1-\rho\Theta^2}\right)$$
$$= -\rho(1-\Theta^2) < 0, \tag{B.52}$$

since $C(1) = 0$, $C(0) = 1$, and $\rho_2^2 = \rho$ when $\alpha = 1$. Further, when $\alpha = 1/2$, $C(\alpha) = C(1-\alpha) = C(\rho_2^2) = \frac{1}{2-\Theta^2}$. Thus,

$$f\left(\frac{1}{2}\right) = (1-\rho)\left(2 - \frac{1}{2-\Theta^2}\right) - \left(1 - \frac{\rho\Theta^2}{2}\right)\left(\frac{2}{2-\Theta^2}\right)$$
$$= \frac{1 - 2\Theta^2 - 3\rho(1-\Theta^2)}{2-\Theta^2}. \tag{B.53}$$

It is clear that $f(1/2) < 0$ if $\Theta^2 \geq 1/2$. Notice that $f(1/2) < 0$ is also true if

$$\rho \geq \frac{1-2\Theta^2}{3(1-\Theta^2)}. \tag{B.54}$$

The right hand side of (B.54) is decreasing in Θ^2, and as Θ^2 goes to zero, the right hand side of (B.54) goes to 1/3. Thus, $\rho \geq 1/3 \Rightarrow f(1/2) < 0$. Therefore, if $f(\alpha)$ is convex on $[1/2, 1]$, then $f(\alpha) < 0$ for all $\alpha \in [1/2, 1]$. From (B.51), one can verify that

$$f''(\alpha) = -(1-\rho)C''(\alpha) + 2\rho\Theta^2\left(C'(1-\alpha) + C'(\rho_2^2)\right) - (1-\rho\alpha\Theta^2)\left(C''(1-\alpha) + C''(\rho_2^2)\right)$$
$$= \frac{2(1-\rho)\Theta^2(1-\Theta^2)}{(1-\alpha\Theta^2)^3} + \frac{2\rho\Theta^2(1-\Theta^2)}{(1-(1-\alpha)\Theta^2)^2} + \frac{2\rho(1-2\rho)\Theta^2(1-\Theta^2)}{(1-\rho_2^2)^2}$$
$$+ \frac{2(1-\rho\alpha\Theta^2)\Theta^2(1-\Theta^2)}{(1-(1-\alpha)\Theta^2)^3} + \frac{2(1-\rho\alpha\Theta^2)(1-2\rho)^2\Theta^2(1-\Theta^2)}{(1-\rho_2^2\Theta^2)^3}, \tag{B.55}$$

where the differentiation is with respect to α. Examining (B.55), it is clear that $f''(\alpha) > 0$

for $\rho \leq 1/2$. Now, suppose $\rho > 1/2$. Notice that

$$\frac{2\rho(1-2\rho)\Theta^2(1-\Theta^2)}{(1-\rho_2^2)^2} + \frac{2(1-\rho\alpha\Theta^2)(1-2\rho)^2\Theta^2(1-\Theta^2)}{(1-\rho_2^2\Theta^2)^3} = \frac{2(1-2\rho)(1-\rho)(1-\rho\Theta^2)\Theta^2(1-\Theta^2)}{(1-\rho_2^2)^3}. \quad \text{(B.56)}$$

Now,

$$\frac{2(1-\rho)\Theta^2(1-\Theta^2)}{(1-\alpha\Theta^2)^3} - \frac{2(2\rho-1)(1-\rho)(1-\rho\Theta^2)\Theta^2(1-\Theta^2)}{(1-\rho_2^2)^3} = \frac{2(1-\rho)\Theta^2(1-\Theta^2)}{(1-\alpha\Theta^2)^3(1-\rho_2^2\Theta^2)^3}\left[(1-\rho_2^2\Theta^2)^3 - (1-\alpha\Theta^2)^3(2\rho-1)(1-\rho\Theta^2)\right]. \quad \text{(B.57)}$$

Notice that

$$(1-\rho_2^2\Theta^2)^3 \geq (1-\alpha\Theta^2)^3 > (1-\alpha\Theta^2)^3(2\rho-1)(1-\rho\Theta^2), \quad \text{(B.58)}$$

where the first inequality is true since $\alpha \geq \rho_2^2$ for all $\alpha \in [1/2, 1]$, and the second inequality holds because $(2\rho-1)(1-\rho\Theta^2) < 1$. Therefore, $f''(\alpha) > 0$ for all ρ ($f(\alpha)$ is convex). Thus, if $\Theta^2 \geq 1/2$ or $\rho \geq 1/3$, then $A < 0$ for all $\alpha \in [1/2, 1]$, which implies that $\overline{y}_1 \leq \overline{y}^c$ and $\underline{y}_1 \geq \underline{y}^c$. The principal learns less about the agent's first-period private information than she would by setting the first period output targets equal to the commitment optimum. □

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Susan Athey and Ilya Segal. An efficient dynamic mechanism. *Econometrica*, 81:2463–2485, 2013.

[2] David P. Baron and David Besanko. Regulation and information in a continuing relationship. *Information and Economic Policy*, 1:267–302, 1984.

[3] Marco Battaglini. Long term contracting with markovian consumers. *American Economic Review*, 95:637–658, 2005.

[4] Marco Battaglini. Optimality and renegotiation in dynamic contracting. *Games and Economic Behavior*, 60:213–246, 2007.

[5] Marco Battaglini and Stephen Coate. Pareto efficient income taxation with stochastic abilities. *Journal of Public Economics*, 92:844 – 868, 2008.

[6] Gary S. Becker. Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70:9–49, 1962.

[7] Dirk Bergemann and Juuso Valimaki. Dynamic mechanism design: An introduction. Discussion paper no. 3002, Cowles Foundation for Research in Economics, Yale University, 2017.

[8] Joseph S. Berliner. *Factory and manager in the USSR*. Harvard University Press, Cambridge, Mass., 1957.

[9] David Blackwell. Comparison of experiments. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1951.

[10] B. Caillaud, R. Guesnerie, and P. Rey. Noisy observation in adverse selection models. *Review of Economic Studies*, 59:595–615, 1992.

[11] Eric Cardella and Briggs Depew. Output restriction and the ratchet effect: Evidence from a real-effort work task. *Games and Economic Behavior*, 107:182–202, 2018.

[12] H. Lorne Carmichael and W. Bentley MacLeod. Worker cooperation and the ratchet effect. *Journal of Labor Economics*, 18:1–19, 2000.

[13] Gary Charness, Peter Kuhn, and Marie Claire Villeval. Competition and the ratchet effect. *Journal of Labor Economics*, 29:513–547, 2011.

[14] Jay Pil Choi and Marcel Thum. The dynamics of corruption with the ratchet effect. *Journal of Public Economics*, 87:427–443, 2003.

[15] Daniel Clawson. *Bureaucracy and the labor process: The transformation of U.S. industry, 1860-1920*. Monthly Review Press, New York, 1980.

[16] Pascal Courty and Hao Li. Sequential screening. *Review of Economic Studies*, 67:697–717, 2000.

[17] Rahul Deb and Maher Said. Dynamic screening with limited commitment. *Journal of Economic Theory*, 159:891–928, 2015.

[18] Mats Dillen and Michael Lundholm. Dynamic income taxation, redistribution, and the ratchet effect. *Journal of Public Economics*, 59:69–93, 1996.

[19] Xavier Freixas, Roger Guesnerie, and Jean Tirole. Planning under incomplete information and the ratchet effect. *Review of Economic Studies*, 52:173–191, 1985.

[20] Dino Gerardi and Lucas Maestri. Dynamic contracting with limited commitment and the ratchet effect. Working paper, 2017.

[21] Robert Gibbons. Piece-rate incentive schemes. *Journal of Labor Economics*, 5:413–429, 1987.

[22] Bengt Holmstrom. Moral hazard and observability. *The Bell Journal of Economics*, 10:74–91, 1979.

[23] Thomas D. Jeitschko and Leonard J. Mirman. Information and experimentation in short-term contracting. *Economic Theory*, 19:311–331, 2002.

[24] Thomas D. Jeitschko, Leonard J. Mirman, and Egas Salgueiro. The simple analytics of information and experimentation in dynamic agency. *Economic Theory*, 19:549–570, 2002.

[25] Thomas D. Jeitschko and John A. Withers. Dynamic regulation with stochastic costs: Signal dampening, experimentation, and the ratchet effect. Working paper, 2018.

[26] Jean-Jacques Laffont and Jean Tirole. Using cost observation to regulate firms. *Journal of Political Economy*, 94:614–641, 1986.

[27] Jean-Jacques Laffont and Jean Tirole. Comparative statics of the optimal dynamic incentive contract. *European Economic Review*, 31:901–926, 1987.

[28] Jean-Jacques Laffont and Jean Tirole. The dynamics of incentive contracts. *Econometrica*, 56:1153–1175, 1988.

[29] Jean-Jacques Laffont and Jean Tirole. *A theory of incentives in procurement and regulation*. MIT Press, 1993.

[30] Jean-Jacques Laffont and Jean Tirole. Pollution permits and compliance strategies. *Journal of Public Economics*, 62:85–125, 1996.

[31] Hugh Macartney. The dynamic effects of educational accountability. *Journal of Labor Economics*, 34:1–28, 2016.

[32] Stanley Matthewson. *Restriction of output among unorganized workers*. Viking Press, New York, 1931.

[33] Nahum D. Melumad and Stefan Reichelstein. Value of communication in agencies. *Journal of Economic Theory*, 47:334–368, 1989.

[34] Leonard J. Mirman, Larry Samuelson, and Edward E. Schlee. Strategic information manipulation in duopolies. *Journal of Economic Theory*, 62:363–384, 1994.

[35] Leonard J. Mirman, Larry Samuelson, and Amparo Urbano. Monopoly experimentation. *International Economic Review*, 34:549–563, 1993.

[36] James Mirrlees. The optimal structure of incentives and authority within an organization. *The Bell Journal of Economics*, 7:105–131, 1976.

[37] David Montgomery. *Workers' control in America: Studies in the history of work, technology, and labor struggles*. Cambridge University Press, New York, 1979.

[38] Alessandro Pavan, Ilya Segal, and Juuso Toikka. Dynamic mechanism design: A myersonian approach. *Econometrica*, 82:601–653, 2014.

[39] Pierre Picard. On the design of incentive schemes under moral hazard and adverse selection. *Journal of Public Economics*, 33:305–331, 1987.

[40] Donald Roy. Quota restriction and goldbricking in a machine shop. *American Journal of Sociology*, 57:427–442, 1952.

[41] Vasiliki Skreta. Optimal auction design under non-commitment. *Journal of Economic Theory*, 159:854–890, 2015.

[42] Michael Waldman. Job assignments, signalling, and efficiency. *Rand Journal of Economics*, 15:255–267, 1984.

[43] Martin L. Weitzman. The "ratchet principle" and performance incentives. *Bell Journal of Economics*, pages 302–308, 1980.