Career Concerns And The Optimal Pay-for-Performance Sensitivity

Qi Chen and Wei Jiang*

This draft: February 2006

Abstract

Prior literature suggests that managers’ career concerns provide implicit incentive mechanisms for their behaviors (Fama (1980), Holmstrom (1999)), and may substitute the explicit incentives in their compensation contracts (Gibbons and Murphy (1992)). We show that the substitution effect can be weakened, and even reversed, when managers perform multiple tasks and are concerned about both the level and variability of their reputation. We also find that after taking into consideration managers’ concerns for the variability of their reputation, the optimal pay-for-performance sensitivity can also be increasing in the underlying risk measure. We test these predictions using a large sample of chief executive officers’ compensations over 1993-2003; results are consistent with our model predictions. (JEL: J31; J41; D80)

*The Fuqua School of Business, Duke University, Durham, NC 27708, email: qc2@duke.edu; Columbia Business School, New York, NY 10027, email: wj2006@columbia.edu. We received helpful comments from Jim Anton, Robert Bushman, Harry Evans, Jennifer Francis, Itay Goldstein, Thomas Hemmer, Steve Kaplan, Young Kwon, Tracy Lewis, Mike Riordan, and workshop participants at University of California – Berkeley and HKUST. We are responsible for all remaining errors. The research was supported by the Fuqua School of Business, Duke University and the Eugene Lang Center of Entrepreneurialship, Columbia Business School.
Career Concerns and the Optimal Pay-for-Performance Sensitivity

Abstract

Prior literature suggests that managers’ career concerns provide implicit incentive mechanisms for their behaviors (Fama (1980), Holmstrom (1999)), and may substitute the explicit incentives in their compensation contracts (Gibbons and Murphy (1992)). We show that the substitution effect can be weakened, and even reversed, when managers perform multiple tasks and are concerned about both the level and variability of their reputation. We also find that after taking into consideration managers’ concerns for the variability of their reputation, the optimal pay-for-performance sensitivity can also be increasing in the underlying risk measure. We test these predictions using a large sample of chief executive officers’ compensations over 1993-2003; results are consistent with our model predictions. (JEL: J31; J41; D80)


1 Introduction

This paper examines the relation between managers’ career concerns and the optimal pay-for-performance sensitivities in their compensations. Corporate executives and managers are subject to various incentive mechanisms, including both explicit incentives from their compensation contracts, and implicit incentives from their career concerns, i.e., their concerns for their reputation in the external labor market. A central belief on the role of career concerns, as expressed in Fama (1980) and elaborated in Holmstrom (1999), is that they provide additional monitoring device, and therefore, may substitute explicit incentives. Gibbons and Murphy (1992) further formalize this substitutive relation by showing that in an optimal contract, a manager’s pay-for-performance sensitivity ($PPS$, hereafter) should be lowered when his career concerns are stronger.

However, researchers have also identified other effects of career concerns on managers’ behaviors, some of which are not to shareholders’ benefits. For example, theoretical works by Holmstrom and Ricart i Costa (1986) and Prendergast and Stole (1996) point out the negative effects of career concerns on agents’ behavior. Song and Thakor (2005) show that career concerns of CEOs, interacted with those of the board, distort investment selections and cause weakened governance. Chevalier and Ellison (1999) find empirical evidence that mutual fund managers’ career-concerns may distort their trading behaviors to the detriment of fund shareholders. These imply that career concerns may not be merely a solution to incentive problems, but can also be a source of incentive problems. As a result, the relation between explicit compensation contracts and managers’ career concerns may not be as straight-forward as the substitution effect suggested in prior literature. In this paper, we show that, indeed, the substitutive relation between implicit and explicit incentives can be weakened, and even reversed, depending on the nature of the manager’s career concerns and the type of tasks he performs.

Specifically, we extend Gibbons and Murphy (1992) (GM, hereafter) along two dimensions. First, the manager in our model faces two types of career concerns: reputation concerns (as in GM), which refer to how the manager’s current performance affects the level of his future compensation, and career-risk concerns, which refer to how the manager’s current performance affects the variation in his future compensation. This extension is motivated by the fact that insurance markets for
human capital are likely incomplete (Holmstrom and Ricart i Costa (1986)). As a result, managers are often concerned about the risks associated with their future reputation and compensation.

The second extension in our model is to introduce multiple tasks. In addition to the traditional production effort (as in GM), the manager in our model also exerts an information effort. Unlike production effort, information effort does not directly increase current or future output in the sense of first-order stochastic dominance. Instead, it produces more precise information about the firm’s profitability (which is related to managerial talent), which benefits the firm.

The multitask setting is more descriptive of corporate executives’ responsibilities. Managers, especially Chief Executive Officers (CEOs), are responsible for day-to-day operations, which correspond to production effort in our model. They also spend significant amount of time analyzing firms’ market positions, gathering information about investment opportunities or profitability, and setting future business strategy. These types of activities are captured by information effort in our model. Because these types of effort relate to firms’ future strategy, we use the term strategy effort interchangeably with information effort throughout the paper. We believe both extensions reflect more realistically the multi-dimensional aspects of corporate executives’ responsibilities and incentives (Holmstrom and Milgrom (1991), Milgrom and Roberts (1992)), and the important role that CEOs play in setting corporation strategies in addition to improving production (Dow and Raposo (2005)).

The key insight from our model is that career concerns (more specifically, career-risk concerns) reduce a manager’s incentive for strategy effort. The intuition is the following. While the information uncovered from the manager’s strategy effort helps the firm make better decisions, it also reveals information about the manager’s ability to succeed in the future. More strategy effort provides more precise information to reduce the firm’s ex post uncertainty about the manager’s ability (after the effort is exerted and the information revealed); at the same time, it increases the ex ante uncertainty about how the manager may be perceived by the labor market.1 Such ex ante

1The intuition is the following. For a given level of ex ante uncertainty in a parameter (such as managerial talent), the more precise a signal is, the more the updated assessment of the parameter (after incorporating the signal) deviates from its prior (and hence the more expected risk the manager faces). A more mathematical version of this intuition is in footnote 23.
uncertainty is costly to the manager because it increases the variability of his future compensation. As a result, the manager will under-supply strategy effort, more so when his career-risk concerns are stronger.

We show that to overcome the negative effect of career-risk concerns on the manager’s strategy effort, the firm may find it optimal to increase the manager’s explicit pay-for-performance sensitivity; more so when managers’ career-risk concerns are higher. This weakens the negative relation between PPS and career concerns established in prior literature. Further, because the manager’s career-risk concerns are higher when the underlying uncertainty about his ability is high, the firm may choose a higher PPS in riskier environment. This weakens the negative relation between PPS and project risk from the standard agency model (e.g., Holmstrom and Milgrom (1987)).

In summary, our model generates predictions not suggested by the existing literature. Our model predicts that both the relation between career concerns and PPS and that between risk and PPS are less negative, and may even turn positive, for firms that benefit more from strategy effort (such as firms with high growth potential), or firms where managers’ stakes in firms’ future wealth are small (such as large firms).

These predictions are consistent with anecdotal evidence suggesting that strong implicit incentives and risky environments do not always suppress PPS. For example, managers in start-up firms, or newly appointed CEOs, often receive highly performance-sensitive compensations (such as restricted stocks or stock options), even though they have relatively strong career concerns. Career concerns clearly provide implicit incentives to executives in the information technology (IT) industries, as evidenced by the high turnover rates for IT executives. At the same time, IT executives’ PPS are significantly higher than their counterparts in non-IT firms (Anderson, Banker, and Ravindran (2000), Ittner, Lambert, and Larcker (2003), Murphy (2003)). Kaplan and Stromberg (2004) also find a positive relation between PPS for managers in start-up firms and some measures of risk.

\[^2\]Empirical support for the negative risk-incentive relation is mixed (see Demsetz and Lehn (1985), Jensen and Murphy (1990) and Aggarwal and Samwick (1999) for examples), and therefore, has prompted several authors to propose alternative explanations (e.g., Core and Qian (2001), Prendergast (2002), Hemmer (2002), Raith (2003), and Oyer and Schaefer (2005)). See also Chiappori and Salanie (2003) and Prendergast (2002) for recent surveys.

of external risk. To the extent that strategy effort is important in these industries, our model offers an explanation for the above seemingly counter-intuitive facts.

We test our model predictions using data on CEO compensation from the ExecuComp database from 1993-2003. We find that consistent with GM, CEO tenure and age (as inverse proxies for the strength of career concerns) are both positively correlated with PPS in CEO contracts. However, more relevant to our model predictions, we find that the relations are less positive for firms with higher growth potential and for larger firms. Similarly, we also find that while the average relation between PPS and firm risk is negative, they are less negative for large and high growth firms. These results are consistent with our model predictions.

The SEC’s decision in early 2006 to amend disclosure requirements for executive and director compensation reflects a growing attention from investors and regulators on executive pays. Our paper makes at least two contributions to the literature on executive compensation. First, it complements the theoretical literature on the relation between explicit and implicit incentives, and on the determinants of pay-for-performance sensitivity. In particular, we analyze how different types of implicit incentives jointly affect PPS. Our model endogenizes the information structure in the standard reputation concerns model. In our model compensation contracts affect the manager’s information effort choice, which, in turn, affects the information available to the principal. Our paper differs from Meyer and Vickers (1997), Milbourn, Shockley, and Thakor (2001), Milbourn (2003), and Goel, Nanda, and Narayanan (2004) in that our manager performs multiple tasks that affect both the level and the variability of his expected future compensation.5 Our paper shares a similar feature with Barron and Waddell (2003) in that the agent exerts non-contractible information effort. However, Barron and Waddell (2003) study a one-period model where the agent

4 Because these risk measures capture the external risk about which venture capitalists and managers are symmetrically informed, Kaplan and Stromberg (2004) conclude that this positive relation is hard to reconcile with a standard agency model (e.g., Holmstrom and Milgrom (1987)) or an information asymmetry explanation (e.g., Pendergast (2002)). In our model, the principal and the agent have incomplete but symmetric information about the underlying project profitability.

5 Dewatripont, Jewitt, and Tirole (1999) also study career concerns with multiple tasks. But their multiple tasks are of the same nature (i.e., they all affect current-period output in the first-order stochastic dominance manner). Further, they do not consider explicit contracts.
exerts only information effort and privately observes the outcome from this effort. In contrast, we
analyze a dynamic model where the outcome of the agent’s effort is observed by both the agent
and the principal. The observability of the outcome of the agent’s effort exacerbates the agent’s
career-risk concerns.

Our paper also adds to the empirical literature on cross-sectional pay-for-performance hetero-
geney in executive compensations. We find large-sample evidence supporting the intuitive result
from Fama (1980) that external labor market monitoring partially substitutes for explicit incentive
compensation. More importantly, we document that the degree of this substitution varies with
firm and executive characteristics. Further, to the best of our knowledge, our finding is the first
to identify a link between implicit incentives and the much debated risk-incentive relation.

The rest of the paper is organized as follows. Section 2 sets up the model and provides a lemma
that facilitates solving the model. Section 3 solves the model. Section 4 discusses the testable
hypotheses and presents supporting empirical results. Section 5 concludes.

2 Model set-up

Our model set-up builds on that in Gibbons and Murphy (1992). In setting up the model, we start
with the basic production function and the interpretation of the career-concerns parameter. We
then introduce multitask and career-risk concerns. We next describe the contract form and the
principal’s constraints. We conclude the section by proving a lemma that would facilitate solving
the model.

2.1 Basic production function and career concerns parameter

A risk-neutral principal hires a risk- and effort-averse agent to manage a project that yields sto-

6Gibbons (1998), Prendergast (1999), and Murphy (1999), among others, provide recent surveys of the literature.
the agent’s nonnegative costly effort \( (e_t) \), a productivity measure \( (\pi_t) \), and random noise \( (\varepsilon_t) \):

\[
\begin{align*}
y_1 &= \pi_1 + e_1 + \varepsilon_1, \\
y_2 / k &= \pi_2 + e_2 + \varepsilon_2.
\end{align*}
\]

(1a) (1b)

The noise terms \( \varepsilon_t \) \( (t = 1, 2) \) are assumed to be distributed normal \( N(0, 1/p_\varepsilon) \),\(^7\) mutually independent, and independent of \( \pi_t \). The coefficient \( k \) in (1b) measures the scale of production in the second period relative to the first period and captures the firm’s (exogenous) growth potential. In a multiple-period context, \( y_2 \) can be viewed as the sum of outputs from all future periods.

The project’s profitability \( \pi_t \) is an exogenous random variable that correlates over time through the following AR(1) system:

\[
\begin{align*}
\pi_1 &= \theta, \\
\pi_2 &= \rho \theta + \sqrt{1 - \rho^2} \varepsilon = \rho \pi_1 + \sqrt{1 - \rho^2} \varepsilon,
\end{align*}
\]

(2a) (2b)

where \( \theta \) is the part of the agent’s ability or talent that persists over time. As is standard in the literature, we assume that the principal and the agent do not observe the true \( \theta \), but share the common prior that \( \theta \) is distributed normal \( N(0, 1/p_\theta) \).\(^8\) Normalizing the mean of \( \theta \) to zero is without loss of generality as we can view the value of \( \theta \) as its deviation from the unconditional mean. In equation (2b), \( \varepsilon \) is a zero-mean normal innovation term in the second period’s profitability \( (\pi_2) \) and is independent of \( \pi_1 \). We normalize the variance of \( \varepsilon \) to be the same as that of \( \theta \). This way, \( \pi_1 \) and \( \pi_2 \) have the same total unconditional variance for all values of \( \rho \). As will be clear later, both \( \rho \) and the total ex ante uncertainty about \( \pi_t \) affect the agent’s incentives. This normalization allows us to isolate the effect of \( \rho \), our measure of the intensity of the agent’s career concerns, as discussed next.

\(^7\)Throughout the paper, \( \sigma_x^2 \) denotes the variance of a random variable \( x \), and \( p_x \equiv 1/\sigma_x^2 \) denotes the inverse of the variance (i.e., the precision).

\(^8\)An intuitive interpretation from Prendergast and Topel (1996) is that \( \theta \) measures the fit between the agent and his job, and therefore is unknown to both the principal and the agent at the beginning of the project. In the example in the introduction section, \( \theta \) measures the manager’s ability to succeed in the foreign market. Laffont and Tirole (1988) and Lewis and Sappington (1997), among others, discuss the optimal contract when the agent has pre-contracting private information about \( \theta \).
The parameter $\rho \in [0, 1]$ in equations (2a) and (2b) measures the project profitability persistence that is due to the persistence in the agent’s ability. It is interpreted as our measure for the intensity of the agent’s career concerns, for two main reasons. First, $\rho$ captures the sensitivity of the agent’s future compensation to current performance. Both current and potential employers will rely on the first-period signals to update their assessment of the agent’s value in the second period. The higher the value of $\rho$ is, the more the employers rely on first-period signals to assess the agent’s ability. Thus, higher $\rho$ implies higher sensitivity of the agent’s future compensation to current-period performance. As a comparison, most studies on career concerns, such as GM, Holmstrom (1999), and Meyer and Vickers (1997), assume $\rho$ to be one, which is a special case in our set-up.

As equations (2a) and (2b) indicate, future profitability replicates current profitability without any innovation when $\rho = 1$. As a result, the agent’s future compensation is very sensitive to his current performance. In the other extreme when $\rho = 0$, current performance is not at all informative about future productivity, in which case the agent has no career concerns and the dynamic contracting is reduced to a repeated single-period contracting. We allow $\rho$ to vary between 0 and 1 to facilitate our analysis of how the explicit pay-for-performance sensitivity varies with the degree of the agent’s career concerns.

The second reason is that $\rho$ captures the tenure effect. Recall that we model the second period as a reduced-form representation of all future periods. Thus $\rho$ is likely to correlate positively with the agent’s expected tenure on the job. A shorter expected remaining tenure implies a lower correlation between the agent’s ability and the firm’s future productivity. This point is subtle but important as both age and tenure on the job are standard (inverse) proxies for the intensity of

\footnote{More realistically, firm profitability also depends on factors other than managerial ability. Adding factors to the profitability that are orthogonal to $\theta$ does not affect the results.}

\footnote{To see this point, consider the extreme case when the current manager is replaced after only one period and a new manager is randomly re-drawn from the population of managerial talent with the same unconditional distribution $\mathcal{N}(0, 1/p\theta)$. The distribution of the firm’s profitability will then be re-set and will not be correlated with the distribution if the manager is not replaced. In this case, the current manager’s remaining tenure is zero and so is the correlation between the firm’s current and future profitability. Similar logic implies that the longer the current manager’s expected tenure, the higher the correlation between the firm’s current and future profitability. See Milbourn (2003) for an elaboration on this point.}
career concerns in empirical studies.

2.2 Multitask and career-risk concerns

A novel feature of our model is to recognize that managers’ tasks often involve activities that do not increase the firm’s current output. Rather, these activities relate to the strategic aspects of the firm’s operations and have implications for the firm’s future operations. For example, executives often spend time gathering and analyzing information, investigating market conditions, and/or assessing the firm’s growth potential in foreign markets. To model this aspect of managerial activities, we assume that in the first period, the agent can exert another non-contractible effort, \( a \in [0, 1] \), to produce a publicly verifiable report, \( s \), about the project’s profitability \( (\pi_1) \). We represent the report \( s \) as a noisy signal generated according to

\[
s = \pi_1 + \eta,
\]

where \( \eta \) is a normal noise term independent of \( \pi_1 \) with mean zero and precision \( \frac{a}{1-a}p_\theta \). This particular functional form for the precision of \( \eta \) is for tractability and without loss of generality. It implies that higher \( a \) increases the precision of \( \eta \), hence the informativeness of \( s \) about \( \pi_1 \). In the extreme case of \( a = 1 \), observing \( s \) is equivalent to observing \( \pi_1 \). Because \( \pi_1 \) and \( \pi_2 \) are correlated via managerial ability \( \theta \), \( s \) is informative about the manager’s potential to succeed in the future.\(^{11}\)

To differentiate the agent’s efforts, throughout the paper, we refer to \( a \) as information effort or strategy effort and \( e_t \) \((t = 1, 2)\) in (1a) and (1b) as production effort. Our main model assumes that the principal uses the signal \( s \) for contracting purposes only. In the sensitivity analysis section, we show that introducing additional benefits of \( s \) does not affect our predictions.

To introduce career-risk concerns, we assume that the agent does not have access to a perfect credit market. This assumption is consistent with empirical evidence suggesting that managers cannot perfectly hedge future career risks, especially early in their career (see, e.g., Jin (2002), Garvey and Milbourn (2003)). We assume that the agent’s total utility is time additive.\(^{12}\) The

\(^{11}\) Related to footnote 9, our results do not change if \( s \) is a signal of both managerial talent and other unrelated determinants of profitability. The key is that \( s \) is informative about \( \theta \).

\(^{12}\) That is, the agent’s total lifetime utility is \( U = \sum_{t=1}^{T=2} U(c_t) \), where \( c_t \) is his consumption in period \( t \). The
assumption is made for simplicity and is stronger than needed as it implies that the agent cannot
insure against any of his future incomes. All we need is some insurance incompleteness for the
agent’s reputation or human capital risk.

For tractability, we assume a mean-variance utility function for the agent’s per-period utility.
Thus, the agent’s total expected utility at the beginning of the first period is:

\[ U(w_t; e_t, a) = \left[ E_0(w_1) - c(e_1) - g(a) - \frac{1}{2} Var_0(w_1) \right] + \left[ E_0(w_2) - c(e_2) - \frac{1}{2} Var_0(w_2) \right], \]  
where \( w_t (t = 1, 2) \) is the agent’s salary in period \( t \), \( r \) measures the degree of the agent’s risk
aversion, and \( c(e_t) \) and \( g(a) \) are the costs of production and information effort respectively in
monetary terms. For simplicity, we assume that \( c(e_t) \) is quadratic in \( e_t \) with \( c(e_t) = \frac{1}{2} e_t^2 \), and \( g(a) \)
is convex in \( a \) with \( g(0) = g'(0) = 0 \) and \( g'(1) = \infty \). Throughout the paper, \( E_0(\cdot) \) and \( Var_0(\cdot) \)
represent ex ante (at the beginning of the first period) expectation and variance.

2.3 Contract form and constraints

As in the standard career concerns models, we make two assumptions about feasible contracts.
First, only one-period contracts are enforceable, that is, neither party can commit to ignoring
information revealed after the first period.\(^\text{13}\) Second, in each period \( t \), the contract takes the linear
form of \( w_t = \alpha_t + \beta_t y_t + \lambda_t s \) for some constants \( \alpha_t, \beta_t, \) and \( \lambda_t.\(^\text{14}\) \) Signal \( s \) enters the contract
because it can help the principal filter the output randomness unrelated to the agent’s production
effort. Lack of committed long-term contracts implies that \( \alpha_2, \beta_2, \) and \( \lambda_2 \) can be functions of \( y_1 \)
and \( s \). Given the assumptions, it is instructive and without loss of generality to write the contracts

\[ U = U \left( \sum_{t=1}^{T=2} c_t \right) \] (as in GM) which assumes that a perfect credit market exists for hedging the agent’s future income risks.

\(^\text{13}\)See Gibbons and Murphy (1992) and Meyer and Vickers (1997) for discussions on the equivalence of the op-
timal sequence of one-period contracts and the optimal renegotiation-proof long-term contract. In general, if full
commitment is allowed, career concerns can be perfectly insured ex ante and would be a moot issue.

\(^\text{14}\)Holmstrom and Milgrom (1987) show that linear contracts are a reasonable approximation in the context of CARA
utility and normal distributions. See Sung (1995), Ou-Yang (2003), and Hemmer (2002) for related discussions on this
approximation. We focus on linear contracts because our primary interest is the relation between pay-for-performance
and career concerns (as opposed to the optimal shape of the contract).
as

\[ w_1 = \alpha_1 + \beta_1 (y_1 - \delta s), \quad (4a) \]
\[ w_2 = \alpha_2 (s, y_1) + \beta_2 (s, y_1) y_2, \quad (4b) \]

where \( \delta = -\lambda_1 / \beta_1 \), and both \( \alpha_2 \) and \( \beta_2 \) are linear functions of \( y_1 \) and \( s \).\(^{15}\)

In setting the contracts, the principal observes three types of constraints. The first is the incentive compatibility constraint which states that the agent will choose \( e_t \) and \( a \) to maximize his expected utility for any given contract. The second is the time consistency constraint which states that, at the start of period two, the principal will use the information gained from period one to set the period two contract. The third is the agent’s participation constraints for both periods, which we discuss next.

Participation constraint ensures that the agent’s expected utility from entering the contract is as least as much as his best outside option at the time of the contract (denoted \( U_t \)). The agent’s outside option value at the beginning of period one \( (U_1) \) is exogenous and is normalized to zero. The agent’s period two reservation utility \( (U_2) \) is endogenous and depends on the information revealed from period one. If the period one information reflects favorably on his ability, the manager’s outside option value (and hence his reservation level) will rise. Following Meyer and Vickers (1997), we assume that \( U_2 \) is proportional to the expected total surplus in period two. That is, \( U_2 = b \cdot TS_2 \) where \( TS_2 \) is the expected second-period output net of the agent’s effort and risk-premium cost and \( b \in (0, 1) \) measures the agent’s share.\(^{16}\)

A positive \( b \) is necessary for career concerns to affect the agent’s behavior in period one. Most prior studies assume \( b = 1 \), which is equivalent to assuming perfectly competitive and risk neutral firms with identical technologies. Allowing \( b \) to vary between 0 and 1 is a more general assumption. \( b \) can be interpreted as the bargaining power of the agent vis-a-vis the principal. As long as both

\(^{15}\)These are the standard results with normal distributions and linear contracts. See Meyer and Vickers (1997) for examples and detailed derivations.

\(^{16}\)The alternative specification is that the manager’s outside opportunity is only proportional to his perceived ability, but not to the second-period pure contracting surplus. This specification does not change the qualitative results.
the principal and the agent are better off continuing the relationship than discontinuing it, any Nash-bargaining would result in a strict interior split of the total surplus, i.e., $0 < b < 1$. The agent’s bargaining power can come from his ability to threaten to take the project (and the surplus) to another principal. Carrying out such a threat is not without cost, e.g., the cost of job searching and relocation, and the cost of losing firm-specific human capital. The parameter $b$ can also be affected by the firm’s cost to replace the manager. It is crucial to our model that the agent captures some, but not all, of the surplus attributable to his ability.

We conclude the set-up by summarizing the timeline in Figure 1:

<table>
<thead>
<tr>
<th>$t = 0$</th>
<th>$0 &lt; t &lt; 1$</th>
<th>$t = 1$</th>
<th>$1 &lt; t &lt; 2$</th>
<th>$t = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>The agent chooses both $e_1$ and $a$ efforts.</td>
<td>$y_1$ and $s$ observed; 1st-period contract executed; and 2nd-period contract signed.</td>
<td>The agent chooses $e_2$ effort.</td>
<td>$y_2$ observed and 2nd-period contract executed.</td>
<td></td>
</tr>
<tr>
<td>1st-period contract signed (i.e., wage payment contingent on $y_1$ and $s$ agreed).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Timeline

2.4 Contract on signal $s$

Since $\pi_t$ is exogenously determined, the optimal performance metric should be the output filtered out, as much as possible, of the influence of $\pi_t$. If the precision of $s$ is exogenously given, Myers and Vickers (1997) show that the optimal weight on $s$, $\delta$ in (4a), should be set at $\delta^* = a$. However, in our model the precision of $s$ is affected by the agent’s information effort $a$, which is not contractible. Further, the contract is offered to the agent before he chooses $a$. Thus, the weight on $s$ cannot

---

17 In its simplest form, Nash bargaining maximizes the product of the two parties’ gains over their respective outside opportunities. Any interior split of rent ($b \in (0, 1)$) yields a positive product, which dominates a corner allocation ($b \in \{0, 1\}$) where the value of the product is zero. This feature is preserved under generalized Nash bargainings.
directly depend on the actual $a$.

We show next that a mechanism exists for the principal to contract on $s$ in the first-period and achieve the same outcome as when $a$ is contractible. Consider a one-period game for the moment. The principal offers the agent the following contract:

$$w_1 = \alpha_1 + \beta_1 (y_1 - \bar{a}s),$$

where $\bar{a}$ is the agent’s announcement of his $a$ effort made after he chooses $a$ but before $s$ is realized, and $\alpha_1$ and $\beta_1$ are constants independent of $\bar{a}$. The following Lemma shows that it is always in the agent’s best interest to announce the true $a$, even when he is not constrained to tell the truth. The proof is in the Appendix (A1).

**Lemma 1** Suppose the principal offers the agent a wage schedule in the form of (5). In a one-period model, given any $\alpha_1$ and $\beta_1$, $\bar{a} = a$ is an equilibrium outcome.

Lemma 1 implies that an optimal contract can be contingent on $s$ in a one-period model, even when the precision of $s$ is not contractible. In the two-period set-up, a necessary condition for Lemma 1 to hold is that the announced $\bar{a}$ is used only in the form of (5). For example, the principal cannot use the announcement to punish the agent for not choosing the principal’s desired $a$.\(^{18}\) We assume that $a$ is observable but not contractible.\(^{19}\) Under such circumstances, both the principal and the labor market will use the observed $a$ to determine the second-period compensation for the agent and use the announced $\bar{a}$ only for the first-period contract (and in equilibrium the agent will announce truthfully). This assumption isolates the role of explicit incentives in motivating information effort from their role of extracting information from the agent. It also maintains the symmetric information structure of our model. In the sensitivity analysis section, we show

---

\(^{18}\)In general, the agent’s choice of $a$ is not necessarily the same as what principal would have chosen had $a$ been contractible. Corollary 1 in Sung (1995) provides conditions under which the principal and the agent would choose the same $a$. It is easy to verify that those conditions are satisfied in the one-period version of our model (where $\rho = 0$), but not in a dynamic setting where the second-period incentives depend on the first-period information.

\(^{19}\)Aghion and Bolton (1992) offer a theoretical analysis on observable but not verifiable situations in financial contracting. Hirshleifer, Chordia, and Lim (2001) provide a scenario that motivates this assumption in the specific context of information collection effort.
that full observability of strategy effort \( a \) is not necessary for our results. Our predictions remain unaffected as long as the principal observes some noisy signal for \( a \), and there is some positive correlation between how much the principal relies on the signal \( s \) and the amount of strategy effort exerted by the agent.

### 3 Two-period contracting

#### 3.1 Second-period contract

Using backward induction, we start with the second-period contract. At the beginning of the second period after \( y_1, s \) and \( a \) are observed, the principal chooses \( \alpha_2 \) and \( \beta_2 \) to maximize the expected profit subject to the agent’s participation constraint (IR) and incentive compatible constraint (IC):

\[
\begin{align*}
\max_{\alpha_2, \beta_2} & \quad E(y_2 - w_2(\alpha_2, \beta_2)|s, y_1) \\
\text{s.t.} & \quad E(w_2|s, y_1) - c(e_2) - \frac{r}{2} \text{Var}(w_2|s, y_1) \geq U_2 = b \cdot TS_2, \quad \text{(IR)} \\
& \quad e_2^* \in \arg\max_{e_2} E(w_2|s, y_1) - c(e_2) - \frac{r}{2} \text{Var}(w_2|s, y_1). \quad \text{(IC)}
\end{align*}
\]

Note that the agent chooses only production effort in period two.

Lemma 2 below summarizes the solution to the principal’s second-period problem.

**Lemma 2** The optimal second-period compensation to the agent is as follows:

\[
w_2 = bk\hat{\theta} + \beta_2^*[y_2 - E(y_2|s, y_1)] + bV_2 + \frac{r}{2} \text{Var}(w_2|s, y_1) + c(e_2^*), \quad \text{(6)}
\]

where \( \hat{\theta} = E(\theta|s, y_1) \) is the market’s perception of the agent’s ability after observing \( s \) and \( y_1 \), \( \beta_2^* = \frac{1}{1 + r \text{Var}(\frac{y_2}{w_2}|s, y_1)} \) is the optimal pay-for-performance sensitivity for period two, \( e_2^* = k\beta_2^* \) is the agent’s optimal effort choice, and \( V_2 = \frac{k^2}{2[1 + r \text{Var}(\frac{y_2}{w_2}|s, y_1)]} \) is the contracting surplus from the principal-agent relationship. Further, the marginal effect of information effort \( a \) on \( V_2 \) is positive and is increasing in \( k \) (i.e., \( \frac{\partial V_2}{\partial a} > 0 \) and \( \frac{\partial^2 V_2}{\partial a \partial k} > 0 \)).

The proof for Lemma 2 follows standard procedures and is delegated to the Appendix (A2). Lemma 2 offers two insights: The first is the decomposition of the second period contact in the form
of (6). Each term in (6) captures a different source of the agent’s future compensation. The first term \(bk\rho\theta\) represents the compensation for the agent’s perceived ability, that is, the agent receives a share \(b\) of the surplus that is attributed to his ability. The second term \(\beta^2_2[y_2 - E(y_2|s,y_1)]\) is the agent’s payment that is explicitly tied to his performance in excess of the benchmark \(E(y_2|s,y_1)\). It reflects the fact that an optimal contract filters out the randomness unrelated to the agent’s effort. The third term is the agent’s share \(b\) of the second-period contracting surplus \(V_2\). The final two terms are compensations for the agent’s risk premium and his cost of effort.

The second insight from Lemma 2 is that information effort \(a\) increases the second-period contracting surplus \(V_2\). The intuition for this result is that higher \(a\) makes signal \(s\) more informative about the underlying profitability \(\pi_1\), and hence \(\pi_2\) by its correlation with \(\pi_1\). A more informative \(s\) helps the principal filter out non-effort related noise in the second-period output \(y_2\). This reduces the conditional risk premium required by the agent and benefits the firm. The benefits are higher when the firm’s growth potential becomes larger (i.e., higher \(k\)).

### 3.2 First-period contract

Back to the first period, the agent chooses production effort \(e_1\) and information effort \(a\) to maximize his expected total utility, taking into account their impact on his compensations in both periods \(w_1\) and \(w_2\), as in (5) and (6), respectively):

\[
\max_{e_1, a} \left[ E_0(w_1) - c(e_1) - g(a) - \frac{r}{2} Var_0(w_1) \right] + \left[ E_0(w_2) - \frac{r}{2} Var_0(w_2) - c(e_2) \right].
\]

Omitting terms unrelated to the agent’s choice variables, we can rewrite the agent’s objective function as (see the Appendix (A3) for detailed derivation).

\[
\max_{e_1, a} u_0 = E_0(w_1) - c(e_1) + k\rho (b - \beta_2) E_0(\theta) - \frac{r}{2} Var_0(w_1) + bV_2(a) - g(a) - \frac{r}{2} Var_0[E(w_2|s,y_1)].
\]

\(20\) In writing the agent’s objective function this way, we have utilized Lemma 1 which shows that if the agent’s reported information collection effort level \(\tilde{a}\) is only used for the first-period contracting in the form of (5), the agent will report truthfully.
Notice that the last four terms in (7) do not appear in the agents’ objective function in prior studies (such as GM, Meyer and Vickers (1997) and Indjejikian and Nanda (1999)). This is because the agents in their models do not perform strategy effort and therefore cannot affect the variability of their own compensations.

Our main goal is to analyze the relation between the optimal pay-for-performance sensitivity in period one ($\beta_1$) and the degree of the agent’s career concerns ($\rho$). We start by examining $\beta_1$ and $\rho$’s effects on the agent’s (production and information) effort choices. First, we note that they both tend to increase the agent’s production effort, as shown in prior studies. It can be easily verified from the agent’s first-order condition with respect to $e_1$:

$$ c' (e_1) = \beta_1 + k\rho (b - \beta_2) \frac{\partial E_0(\bar{\theta})}{\partial e_1} = \beta_1 + \rho kh_y (b - \beta_2), \quad (8) $$

where $h_y$ is the relative precision of the first-period output $y_1$ (see equation (14) in the Appendix (A2)).

Equation (8) indicates that higher explicit incentive ($\beta_1$) leads to higher production effort. Higher reputation concerns ($\rho$) have the same effect if $b - \beta_2 > 0$. To maintain the standard notion of career concerns, we restrict our discussions to the case where $b - \beta_2 > 0$ because this is a necessary condition for reputation concerns to be positive (that is, for the agent to have incentive to strive for a higher, rather than lower, reputation.)\(^{21}\) When reputation concerns are present, a higher $\rho$ implies that a lower $\beta_1$ is needed to induce the same effort $e_1$. This is the main intuition behind the standard substitution effect analyzed in prior literature.

Next, we examine the effects of $\beta_1$ and $\rho$ on the agent’s information effort $a$. In parallel to (8), the agent’s first-order condition with respect to information effort $a$ is:

$$ g' (a) = -r \frac{\partial Var_0(w_1)}{\partial a} + \left[ -r b^2 \rho^2 k^2 \frac{\partial Var_0(\bar{\theta})}{\partial a} \right] + b \frac{\partial V_2}{\partial a} + D, \quad (9) $$

\(^{21}\)GM and most other models assume $b = 1$, therefore, $b > \beta_2$ by assumption. If $0 < b < \beta_2$, then the ratchet effect (i.e., too high a first-period performance raises the benchmark to which the agent’s second-period performance will be compared, thus reducing the agent’s incentive to work hard in the first period) dominates the reputation effect and the net dynamic incentive becomes negative. See Meyer and Vickers (1997) for an elaboration on this point.
where $D = \left[ (b - \beta_2) \frac{\partial h}{\partial a} - h \frac{\partial \beta_2}{\partial a} \right] \rho k (e_1 - \tilde{e}_1)$.\(^{22}\)

The relation between $\beta_1$ on $a$ is positive and is shown in the following lemma (the proof is in the Appendix (A4)).

**Lemma 3** The agent’s incentive for information effort $a$ is increasing in his explicit incentive ($\beta_1$).

The intuition behind Lemma 3 is as follows. Higher $a$ improves the informativeness of the signal $s$, thus helping filter out non-output related noise in the agent’s performance measure. The agent is averse towards his current compensation risk (recall $\text{Var}_0(w_1)$ appears as a cost in the agent’s objective function). Thus, the agent’s incentive for effort $a$ is higher when the marginal effect of $a$ on $\text{Var}_0(w_1)$ is higher. Since this marginal effect is increasing in $\beta_1$ (note $\frac{\partial \text{Var}_0(w_1)}{\partial a}$ is higher in absolute term when $\beta_1$ is higher), the agent’s incentive for $a$ is higher when $\beta_1$ is higher.

Lemma 3 implies that $\beta_1$ carries a dual role of providing explicit incentives for both production and information effort. Whenever information effort $a$ is more valuable, the principal could set a higher $\beta_1$ to induce a higher $a$. However, the effects of $\rho$ on information effort are more complicated, as we discuss next.

The effects of $\rho$ on information effort are captured in terms $B$ and $C$ in (9). Term $B$ shows that the agent’s career concerns dampen his incentive for information effort. This is a key new insight from our model. Term $B$ is the marginal effect of information effort on the ex ante variance of the agent’s future compensation ($\frac{\partial \text{Var}_0[E(w_2|s,y_1)]}{\partial a}$). The Appendix (A5) shows that $\frac{\partial \text{Var}_0[E(w_2|s,y_1)]}{\partial a}$ is positive, and larger in magnitude when $\rho$ is higher. That is, higher information effort is costly to the agent because it increases the agent’s future compensation risk, more so when $\rho$ is higher. This is because higher $a$ increases the informativeness of signal $s$ about profitability $\pi_1$ (hence the agent’s ability $\theta$). Consequently, the posterior estimate of $\theta$ (based on $s$) is more likely to deviate from the prior, increasing the uncertainty about the agent’s future reputation, and therefore, increasing the uncertainty of his future compensation.\(^{23}\) At the same time, higher $a$ does not increase the expected

\(^{22}\)Note that in equilibrium, $e_1 = \tilde{e}_1$, term D disappears and does not affect the agent’s choice of $a$.

\(^{23}\)This follows from the variance identity: $\text{Var}(E(\theta|s)) = \text{Var}(\theta) - E(\text{Var}(\theta|s))$. We know that more informative $s$ reduces the ex post conditional variance $\text{Var}(\theta|s)$. Thus, holding $\text{Var}(\theta)$ constant, $\text{Var}(E(\theta|s))$ will be higher, that is, the manager’s ex ante uncertainty about what the market will think of him is higher.
level of the agent’s reputation (i.e., \( E_0[E(\theta|s)] = E_0(\theta) \) by the law of iterated expectation). As a result, the agent’s career-risk concerns discourage information effort,\(^{24}\) and more so when \( \rho \) is high. This effect works to reduce the supply of information effort from the agent, hence calls for higher explicit incentive for \( a \).

Term \( C \) in (9) indicates that the second-period contracting surplus \( (V_2) \) provides some incentive for the agent to exert information effort (recall that Lemma 2 shows \( \frac{\partial V_2}{\partial a} > 0 \) and the agent’s share \( b \) is positive). When \( \rho \) is high, information effort becomes more useful to improve future surplus. However, this benefit of \( a \) is not fully internalized by the agent because he obtains only \( b < 1 \) share of the future surplus. As a result, he is likely to under-supply information effort, more so when his share of the surplus \( b \) is smaller. This also calls for higher explicit incentive for information effort.

In summary, the agent’s career concerns as measured by \( \rho \) have two opposing effects on the agent’s behaviors, and consequently, two opposing effects on the principal’s optimal \( \beta_1 \) choice. On the one hand, higher \( \rho \) provides stronger incentive (via reputation concerns) for the agent to exert production effort \( e_1 \). This effect tends to lower the optimal \( \beta_1 \), the standard substitution effect as in GM. On the other hand, higher \( \rho \) makes it more important to provide strong incentive for information effort. The reasons, as described above, are that higher \( \rho \) reduces the agent’s incentive to supply information effort (via career-risk concerns), and at the same time, increases the demand for information effort on the principal’s side. This effect may lead to a higher \( \beta_1 \). These two opposing effects make the relation between \( \beta_1 \) and \( \rho \) ambiguous. The following proposition formalizes this relation. The proof is in the Appendix (A6).

**Proposition 1** Let \( \beta_1^* \) be the optimal incentive \( \beta_1 \) for the principal. The substitutability of explicit and implicit incentives is weaker with the presence of career-risk concerns, and more so when the growth potential \( (k) \) is higher (i.e., \( \partial \left( \frac{d\beta_1^*}{d\rho} \right) / \partial k > 0 \)). Further, \( \beta_1^* \) is increasing in \( \rho \) (i.e., \( \frac{d\beta_1^*}{d\rho} > 0 \), or the substitutive relation between implicit and explicit incentive is reversed), when \( k \) is sufficiently large.

The intuition for Proposition 1 is as follows. We have shown that higher \( \beta_1 \) increases both

\(^{24}\)This effect is similar to Goldman (2004) where a manager is reluctant to make an investment that makes stock price more informative about his performance.
information effort \( a \) and production effort \( e_1 \) \((\frac{\partial a}{\partial \beta_1} > 0 \text{ and } \frac{\partial e_1}{\partial \beta_1} > 0)\), while higher \( \rho \) increases production effort but decreases information effort \((\frac{\partial e_1}{\partial \rho} > 0 \text{ and } \frac{\partial a}{\partial \rho} < 0)\). When \( k \) is low, information effort is not very important. Therefore, as \( \rho \) increases, \( \beta_1 \) can be lowered to induce the same level of production effort. When \( k \) is high, in contrast, information effort is relatively more important. As a result, when \( \rho \) is higher, \( \beta_1 \) needs to increase to induce the optimal level of information effort \( a \). When \( k \) is sufficiently high, this positive effect may dominate the substitution effect between \( \beta_1 \) and \( \rho \).

The next Proposition establishes how, in the presence of career-risk concerns, \( \beta_1^* \) relates to another parameter of interest, \( \sigma_\theta^2 \). \( \sigma_\theta^2 \) measures the ex ante uncertainty about the agent’s ability and the firm’s profitability. When the agent cannot choose strategy effort \( a \), the optimal \( \beta_1 \) is decreasing in \( \sigma_\theta^2 \) for two reasons. First is the standard risk-incentive trade-off story: high \( \sigma_\theta^2 \) increases the risk premium that the agent demands for accepting a contract based on stochastic output, thus lowers the optimal \( \beta_1 \). The second reason is that \( \sigma_\theta^2 \) increases the implicit incentive from reputation concerns (recall that \( \frac{\partial E_0(\hat{\theta})}{\partial e_1} \) is increases in \( \sigma_\theta^2 \)), which, in turn, reduces the need for explicit incentive.

When the agent chooses information effort \( a \), \( \sigma_\theta^2 \) has two additional impacts. The first is to increase \( a \)’s marginal impact on the second-period contracting surplus (i.e., \( \frac{\partial V_2}{\partial a} \) is positive and increasing in \( \sigma_\theta^2 \)). This increases the benefit of, and hence the demand for, \( a \), which works to increase \( \beta_1 \). The second impact is to exacerbate the negative effect of career-risk concerns on the agent’s information effort (recall that \( \frac{\partial \text{Var}_0(\hat{\theta})}{\partial \beta_1} \) is positive and increasing in \( \sigma_\theta^2 \)). This effect dampens the agent’s incentive for, and hence reduces the supply for, \( a \), similar to the effect of \( \rho \) as in Proposition 1. As a result, the principal may need to increase the optimal \( \beta_1 \). Together, when the two positive effects dominate, the optimal \( \beta_1 \) can be positively related to \( \sigma_\theta^2 \). The next Proposition formalizes this idea. The proof is in the Appendix (A7).

**Proposition 2**  
The negative relation between risk \( (\sigma_\theta^2) \) and incentive \( (\beta_1) \) is weaker with the presence of career-risk concerns, and more so when the growth potential \( (k) \) and the risk level \( (\sigma_\theta^2) \) are higher. Further, \( \beta_1^* \) is increasing in \( \sigma_\theta^2 \) when \( k \) and \( \sigma_\theta^2 \) are sufficiently large.
A corollary from Propositions 1 and 2 is that the positive relation between $\beta_1$ and $\rho$ (or $\sigma_\theta^2$) is, other things equal, more likely to hold when the agent’s share of the total surplus, $b$, is lower. This is because a lower $b$ widens the divergence between the principal and the agent’s objectives, reducing the agent’s incentive to exert information effort. As a result, the principal has to impose a higher-powered incentive in period one to encourage more information effort.

### 3.3 Extension and sensitivity analyses

#### 3.3.1 Other benefits of information effort

So far, our analysis has assumed that information effort $\alpha$ is used only for contracting purposes. We model $\alpha$’s benefits in terms of its effect on the second-period contracting surplus $V_2$. However, it is more reasonable to assume that information effort has additional benefits. These benefits may stem from, for example, optimal allocation of the firm’s resources or improved planning of investments in uncertain markets. As long as the marginal impact of information effort on these benefits are increasing in $\rho$ and $\sigma_\theta^2$ (which is the case for $V_2$),

|\[\partial M_2 / \partial \alpha > 0\]| Intuitively, the higher the $\rho$, the more informative the signal $s$ is for the second period prospects. Thus, $\partial^2 M_2 / \partial a \partial \rho \geq 0$ implies that the marginal benefit of information collection effort is larger when current performance signals have larger correlation with future performance. Similarly, since higher $\alpha$ effort reduces the uncertainty about $\theta$, its marginal benefit is likely to be non-decreasing in the level of ex ante uncertainty (i.e., $\partial^2 M_2 / \partial a \partial \sigma_\theta^2 \geq 0$).

25 These are reasonable assumptions. Denote these additional benefits as $M_2$ with $\partial M_2 / \partial \alpha > 0$. Intuitively, the higher the $\rho$, the more informative the signal $s$ is for the second period prospects. Thus, $\partial^2 M_2 / \partial a \partial \rho \geq 0$ implies that the marginal benefit of information collection effort is larger when current performance signals have larger correlation with future performance. Similarly, since higher $\alpha$ effort reduces the uncertainty about $\theta$, its marginal benefit is likely to be non-decreasing in the level of ex ante uncertainty (i.e., $\partial^2 M_2 / \partial a \partial \sigma_\theta^2 \geq 0$).
3.3.2 Observability of information effort

We are interested in how career-risk concerns affect managers’ strategy effort. For career-risk concerns to arise, it is necessary to have some observability of managers’ strategy effort by the principal and the market. To see this point, suppose neither the principal nor the market observes $a$ and instead forms a belief $\hat{a}$ based on prior information. If the belief $\hat{a}$ does not depend on the actual $a$, then the ex ante uncertainty about the agent’s perceived ability ($\text{Var}_0(\hat{\theta})$) is actually decreasing in $a$ for given $\hat{a}$ because $a$ lowers the unconditional variance of $s$. As a result, there will be no career-risk concerns effect.\textsuperscript{26}

A more realistic assumption is that the principal does not directly observe $a$, but observes a noisy signal about $a$: $\omega = a + \varepsilon_{\omega}$, where $\varepsilon_{\omega} \sim \mathcal{N}(0, 1/p_\omega)$. This assumption introduces uncertainty into the precision of a random variable ($s$), which, in general, does not allow a closed-form solution (see Subramanyam (1996) for a discussion on uncertain precision). However, our results go through under the following two conditions (details shown in the Appendix A8). The first condition is that $\omega$ does not directly depend on the realized value of $s$. This implies that greater information effort by the agent does not lead to a more favorable signal (it does increase the precision of the signal). This condition is necessary to preserve the linear additive structure of Bayesian updating with uncertain precision. The second condition is that the agent does not control the precision of $\omega$ ($p_\omega$). That is, the agent has no control over how his information effort is monitored.

To summarize, the mechanism of career-risk concerns does not require full observability of the agent’s strategy effort $a$, but requires a positive link between $a$ and the principal’s assigned weight on signal $s$. This link builds a positive connection between $a$ and the variability of the agent’s reputation.

\textsuperscript{26}To see this, note that $\hat{\theta} = E(\theta | s, y_1, \hat{a}) = \hat{h}_y (y_1 - \hat{e}_1) + \hat{h}_s s$ (here $\hat{h}_y$ and $\hat{h}_s$ are analogously defined as in (14) in the Appendix except that $a$ is replaced by $\hat{a}$). Straight-forward algebra shows that holding $\hat{a}$ constant, $\text{Var}_0(\hat{\theta})$ is decreasing in $a$ because $a$ only lowers the unconditional variance of $s$. An equilibrium will still exist because the $a$ effort is bounded between $[0, 1]$ with zero (infinite) marginal cost at the low (high) end, but the result will follow the standard multi-task effort-motivation model, rather than the career-risk concerns mechanism we identify in this paper.
4 Empirical Evidence

In this section, we test our model predictions using data on the compensation of chief executive officers (CEOs) in U.S. public companies. CEO compensation is suitable for our tests because a CEO’s responsibilities involve multiple tasks and career concerns are an important factor in their incentive structure; both multitask and career concerns are key elements in our model.

4.1 Methodology and Testable Hypotheses

Our empirical analysis is based on the standard specification employed in the literature. The standard specification is to estimate the following equation (e.g., Aggarwal and Samwick (1999), Milbourn (2003)):

\[ w_{i,j,t} = \alpha_0 + \beta \cdot y_{i,j,t} + \lambda \cdot Risk_{i,j,t} \cdot y_{i,j,t} + \gamma \cdot Career_{i,j,t} \cdot y_{i,j,t} + Control + \epsilon_{i,j,t}, \]  

(10)

where \( w_{i,j,t} \) is the compensation for CEO \( i \) for firm \( j \) in year \( t \), \( y_{i,j,t} \) is a performance measure (such as change in shareholders’ wealth), \( Risk_{i,j,t} \) is a measure of the firm’s underlying risk, \( Career_{i,j,t} \) is a proxy for the strength of the CEO’s career concerns, and \( Control \) is a set of control variables (e.g., firm size). For notational brevity, we omit subscripts \( i, j, \) and \( t \) in subsequent equations.

The average pay-for-performance sensitivity (\( PPS \)) is estimated with the coefficient \( \beta \). The coefficients \( \lambda \) and \( \gamma \) capture how \( \beta \) (\( PPS \)) varies cross-sectionally with firm risk and managers’ career concerns. For example, using compensation data from ExecuComp, Aggarwal and Samwick (1999) document a negative \( \lambda \), suggesting a negative relation between \( PPS \) and risk measures. Gibbons and Murphy (1992) find that \( \beta \) is higher for older CEOs or CEOs near retirement, suggesting a negative \( \gamma \), which is consistent with their prediction that, on average, career concerns substitute for \( PPS \).

Our model predictions are about the cross-sectional determinants of \( \gamma \) and \( \lambda \). Specifically, our model implies that while the average \( \gamma \) and \( \lambda \) in (10) can be negative, they will be less negative, or even turn positive, for firms with high growth potential (i.e., \( k \) or \( Growth \)) and for firms whose CEOs have relatively small stakes in firms’ total wealth (i.e., \( b \) or \( Share \)). Therefore, the empirical implications from our analysis are \( \partial \gamma / \partial Growth > 0 \), \( \partial \gamma / \partial Share < 0 \), \( \partial \lambda / \partial Growth > 0 \) and
\[ \partial \lambda / \partial \text{Share} < 0. \]

We use the same variables in Gibbons and Murphy (1992) to capture the strength of CEOs’ career concerns (Career): the CEO’s age (Age) and the number of years since the executive became the firm’s CEO (Tenure). Theory suggests that both Age and Tenure are negatively related to the strength of CEOs’ career concerns. This is because older CEOs have shorter expected time horizon remaining to capture the benefits of a good reputation; and because there is less uncertainty about CEOs’ ability as their tenure grows. Given that we only have information about the CEO’s tenure with his current job, Age should provide more information about how far along the CEO is into his career. In addition, as a robustness check, we also use the first principal component of the two measures (PC(Career)) as a summary measure for Career.\(^{27}\) We expect a positive relation between PPS and Tenure, Age and PC(Career), but less positive for high growth firms and for firms in which CEOs capture a lower share.

We use two proxies for CEOs’ shares in firms’ total wealth: firms’ sales revenue in year \(t - 1\) (Sales) or market capitalization (MV) at the beginning of year \(t\). They are inverse proxies, with larger values indicating smaller CEOs’ stakes in the firms’ wealth. Morck, Shleifer, and Vishny (1988) find that managers own fewer percentages of shares in large corporations. Similar evidence is documented by others, e.g., Jensen and Murphy (1990), and Core and Guay (2002).\(^{28}\)

We use three proxies for a firm’s growth potential (Growth): the first is the firm’s annual growth rate of sales revenue (Grow_sales) over the three years prior to year \(t\), the second is the firm’s market-to-book equity ratio (M/B) at the beginning of year \(t\), and the third is the negative log value of the firm’s age (Firmage), defined as the number of years since the firm’s first appearance on CRSP. Each proxy captures the firm’s growth from a different perspective. Grow_sales measures historical growth; M/B captures future growth potential as impounded in the forward-looking stock.

\(^{27}\)A principal component is a linear combination of the component variables, with coefficients equal to the eigenvectors of the correlation matrix normalized to unit length. The first principal component is associated with the largest eigenvalue. It is a normalized data series that is meant to capture the maximum common variation in the component data series.

\(^{28}\)See Baker and Hall (2004) for a theoretical justification for the negative relation between CEO shares and firm size.
price; and *Firmage* is a direct indicator of the firm’s life cycle stage (young firms tend to have higher growth potential). As a robustness check, we also use the first principal component of all three proxies (*PC*(Growth)) as a summary measure for *Growth*.

We have two proxies for risks and uncertainty. Our model distinguishes between two types of risks: the noise in the performance measure (*Var*(*ε*_t*)) and the uncertainty about project profitability/CEO ability (*Var*(*π*_t*)). Our model predicts that *β* is negatively related to *Var*(*ε*_t*) for all firms, but may be negatively or positively related to *Var*(*π*_t*). Our first risk measure is the return volatility used in the existing literature, measured as the standard deviation of a firm’s monthly dollar return to shareholders over the 48 months prior to year *t* (*$Stdev*$). Since return volatility likely reflects both the variance of noise (*Var*(*ε*_t*)) and the uncertainty about *π*_t* (*Var*(*π*_t*)), our model predicts that *$Stdev*$ is, on average, negatively correlated with *β*, but less so for large or high-growth firms.

Our second risk measure for project profitability/CEO ability is a dummy variable for whether the CEO is hired from the outside (as opposed to promoted from within the firm). Intuitively, we expect that the uncertainty about a CEO’s ability is higher for externally-hired CEOs than for internally-promoted CEOs. If so, Proposition 2 suggests that *PPS* is higher for outside CEOs than for internally-promoted CEOs.29

To summarize, we use the following regression to test our model’s predictions:

\[
\begin{align*}
w & = \alpha_0 + Control + \beta_0 \cdot y + \beta_1 Outside \cdot y + \\
& \quad + \gamma_0 Career \cdot y + \gamma_1 Size \cdot Career \cdot y + \gamma_2 Growth \cdot Career \cdot y + \\
& \quad + \lambda_0$Stdev$ \cdot y + \lambda_1 Size \cdot $Stdev$ \cdot y + \lambda_2 Growth \cdot $Stdev$ \cdot y + \epsilon,
\end{align*}
\]

(11)

where *Career* is either *Tenure* or *Age* and *Growth* is either *Grow_sales*, *M/B*, or *Firmage* (transformed to be the negative log value). The *Control* variables include all the variables interacted with *y* (to control for their average effect on *PPS*) and these variables on their own (to control for their average direct effect on the level of compensation). In all estimations, we also

29 Milbourn (2003) makes a similar prediction under a different setting where he assumes that outside CEOs have higher levels of perceived ability.
include both the firm and year fixed effects to allow different levels of compensation across firms and years. Our main hypotheses are:

- Similar to Gibbons and Murphy (1992), $\gamma_0 > 0$; and as implied by Proposition 1, $\gamma_1 < 0$ and $\gamma_2 < 0$; and

- Similar to Aggarwal and Samwick (1999), $\lambda_0 < 0$; and as implied by Proposition 2, $\lambda_1 > 0$, $\lambda_2 > 0$, and $\beta_1 > 0$.

4.2 Data and Empirical Results

We retrieve data on CEO compensation from Standard and Poor’s ExecuComp database. ExecuComp provides annual compensation data for the five highest paid officers (in terms of salary and bonus) for firms in S&P 500, S&P Midcap 400, and S&P SmallCap 600. We focus on executives’ compensation during their years as CEOs. We identify these years using the BECAMECE and LEFTOFC variables from ExecuComp. BECAMECE indicates the date the executive becomes the CEO and LEFTOFC indicates the date the executive leaves the CEO position.30

We calculate both $y$ and $w$ according to the standard practice in the literature. Specifically, $y$ is the dollar change in a firm’s market equity value during a given year. We calculate a CEO’s total annual compensation $w$ as the sum of the CEO’s flow compensation and value change from the CEO’s holdings of the company’s stocks and options during a given year. The flow compensation measure includes both short-term compensation (e.g., salary, bonus, other annual payments) and long-term compensation (e.g., payouts from long-term incentive plans, the value of restricted stocks granted, the Black-Scholes value of options granted, and other payments such as contributions to benefit plans). Information on CEO age is available from ExecuComp only for one-third of our sample CEOs. For CEOs without age information in ExecuComp, we hand-collect the Age variable from firms’ proxy statements on the SEC’s EDGAR web site. We also exclude 60 CEOs who held more than an average of 25% of their company’s stock over the sample period. We exclude these CEOs because we believe such high ownership causes these CEOs to behave more like an

---

30 We do not use the variable CEOANN to identify the CEO years because ExecuComp often mistakenly carries a CEOANN flag even for years when the executive is no longer the CEO.
owner/principal, rather than an agent. Guay (1999) uses a similar exclusion criterion. Our results are not sensitive if we exclude CEOs with more than 20% or 30% ownership. Our final sample consists of 11,790 firm-year observations, representing 2,954 CEOs for 2,026 firms over 1993-2003, a period spanning both bull and bear markets.

Table 1 presents the definitions and summary statistics for the main variables used in the tests. All variables in dollar terms are deflated (using CPI) to 2001 constant dollars. The median CEO compensation rose from $2.4 million in 1993 to $8.5 million in 2003, after reaching an all-sample low of $0.78 million in 2002 (due to the negative return of the stock market). The all-sample median is $2.8 million. Similar to findings in prior studies, many variables, such as CEO compensation ($w), sales ($Sales), and the dollar risk measure ($$Stdev$$), are highly right-skewed. To mitigate the effect of variable skewness in estimation, we use the log values of $$Stdev$$ and Sales in regressions. (These variables are close to being log-normal.)

Using log values serves the same goal as using rank measures (which is done in some prior studies), but preserves more information about the variables’ actual magnitudes than rank measures. To mitigate the influence of extreme observations, we winsorize all potentially unbounded continuous variables at the 1st and the 99th percentiles (like other recent studies such as Jin (2002) and Garvey and Milbourn (2003)). In addition to OLS regressions, we present results from quantile regressions to mitigate the influence of outlying observations as a robustness check.

To facilitate the interpretation of coefficients on multiplicative regressors, we express non-dummy variables that interact with $y$ as the deviations from their respective sample means. Therefore, the coefficient estimate on $y$ ($\beta_0$) can be readily interpreted as the PPS of a CEO with average characteristics. Similarly, the coefficient estimate on $Career \cdot y$ indicates how PPS for a CEO with average characteristics changes over the CEO’s career.$^{32}$

Table 2 reports our main empirical results. Panel A uses Age and Panel B uses Tenure

---

$^{31}$The transformation is not applicable to $y$ and $w$ because many observations have negative values.

$^{32}$Following equation (11), for a given observation $j$, the marginal effect of $Career$ on PPS is measured by $\gamma_0 + \gamma_1 \cdot (Size_j - \overline{Size}) + \gamma_2 \cdot (Growth_j - \overline{Growth})$, where $\overline{Size}$ and $\overline{Growth}$ are the sample means for Size and Growth. Thus, for a firm with average values of Size and Growth, the effect of $Career$ on its CEO’s PPS is just $\gamma_0$, the coefficient on $Career \cdot y$. 

25
as the proxy variable for Career. Each panel presents specifications using different proxies for Growth or Size. Column 1 is the default specification where Grow_Sales and Ln(Sales) are used. Columns 2, 3 and 5 replaces Grow_Sales with M/B, Firmage and PC(Growth) (the first principal component of Grow_Sales, M/B and Firmage). Column 4 uses MV as firm size proxy. Column 6 uses PC(Career) (the first principal component of Tenure and Age) as the dependent variable (thus Column 6 of Panels A and B are identical). The coefficient estimates on the principal component variables (Columns 5 and 6) are not directly interpretable in terms of economic magnitude. We focus on the signs and significance levels of these coefficients. Finally, Column 7 presents results from quantile regression where the conditional identifying quantile is set to be the sample mean of w. Since all columns obtain overall qualitatively similar results, our discussion below will focus on the default specification in Column 1.

The first several rows in Table 2 establish the following baseline results in the literature. (1) The average magnitude of pay-for-performance sensitivity (PPS). The coefficient estimate for y indicates that the PPS of a CEO with average characteristics is $20 to $25 for every $1,000 change in firm value using linear regressions, and $13 to $15 using quantile regressions. (Throughout the rest of the paper, all PPS are normalized to per $1,000 change in firm value.) These magnitudes are in line with those reported in prior studies (e.g., Hall and Liebman (1998)). (2) The average substitutive relation between implicit and explicit incentives. The coefficients on Age · y in Panel A and on Tenure · y in Panel B are all significantly positive at less than the 1% level, consistent with findings in GM and Core and Guay (1999) that PPS is, on average, higher for older CEOs and CEOs with longer tenure. For example, the coefficient estimate on Age · y from Column 1 in Panel A is 0.35, indicating that PPS increases by $3.5 for every 10 years’ advance in an average CEO’s age. (3) The average negative relations between PPS and firm risk and size. The coefficients for Ln($Stdev) · y and Size · y are significantly negative (at less than the 1% level) across all specifications, confirming prior studies’ finding that PPS is, on average, lower in riskier and larger firms (e.g., Aggarwal and Samwick (1999), Baker and Hall (2004)). For example, if a firm’s sales increases from the average log value of 7.15 to the 75th percentile value of 8.22, its PPS will be lowered by $5.56 (= −5.2 · (8.22 − 7.15)). (4) Higher PPS for outside-hired CEOs. The
coefficient estimate for $Outside \cdot y$ is 4.76 in Column 1 of Panel A (and similar magnitude in other columns), indicating that $PPS$ for outside-hired CEOs is $4.76$ higher than that for internally-promoted CEOs. This result is consistent with our model prediction of a positive relation between $PPS$ and uncertainty about CEO’s ability.

Our main tests focus on the coefficient estimates for the four triple multiplicative terms. First, our model predicts that the rate of substitution between career concerns and explicit incentives should be weaker in large firms and high growth firms. This implies that the coefficients for $Growth \cdot Career \cdot y$ and $Size \cdot Career \cdot y$ should be negative. We find that consistent with this prediction, the coefficient estimates for $Growth \cdot Career \cdot y$ are all negative and significant at 5% or less in Panel A where $Age$ is the proxy for career concerns. The economic magnitude of this coefficient can be interpreted as follows. For a firm with average sales growth rate ($Grow\_Sale = 13\%$), the $PPS$ of its CEO will increase $3.5 for every 10 years’ increase in the CEO’s age. However, for a firm with a sales growth rate at 75th percentile of our sample (i.e., $Grow\_Sale = 20\%$), its CEO’s $PPS$ will increase by only $3.12 (= (0.35 - 0.55 \cdot (0.2 - 0.13)) \cdot 10)$ for every 10 years’ increase in age. Here the coefficient estimate $-0.55$ (on $Growth \cdot Career \cdot y$ in Column 1 of Panel A) suggests that the substitution between career concerns and explicit incentives is weaker in higher growth firms, consistent with our prediction. Using $Tenure$ as a Career proxy yields qualitatively similar results, with overall lower significance level.

Moreover, as predicted by our model, the coefficient estimates on $Size \cdot Career \cdot y$ are negative across all specifications. The coefficients are significant (at less than 1%) using $Tenure$ but less so using $Age$. These results indicate that in large firms where CEOs tend to get smaller shares of firms’ output, the $PPS$ for CEOs are relatively high even when the CEOs are at a relatively early stage in their career.

Second, our model also predicts that the negative relation between risk and explicit incentive is attenuated in large firms and in high-growth firms. Consistent with these predictions, we find that the coefficients for $Growth \cdot Ln\left(\$Stdev\right) \cdot y$ and $Size \cdot Ln\left(\$Stdev\right) \cdot y$ are positive and statistically significant (at less than the 5% level) in most specifications. Using the numbers from the default specification (Column 1 of Panel A), the economic magnitude of these effects
is as follows. For a firm with average sales growth rate \((\text{Grow}_\text{Sale} = 13\%)\), its \(PPS\) will
decrease by $3.28 (=-3.73 \times (5.7 - 4.82)) if the firm’s risk measure increases from the average
value \((\text{Ln}($Stdev) = 4.82)\) to the 75th percentile value \((\text{Ln}($Stdev) = 5.7)\). However, if the
firm’s growth rate is at the 75th percentile value \((\text{Grow}_\text{Sale} = 20\%)\), its \(PPS\) will only decrease
by $2.95 (=[-3.73 + (0.20 - 0.13) \times 5.37] \times (5.70 - 4.82)) for the same change in its risk measure. Similar interpretation applies to the positive coefficient estimate for \(\text{Size} \cdot \text{Ln}($Stdev) \cdot y\). The
\(PPS\) of a firm of average size will decrease by $3.28 if the firm’s risk measure increases from
average to 75th percentile, while the \(PPS\) of a firm of 75th percentile size will decrease only $2.28
(=[-3.73 + (8.22 - 7.15) \times 1.06] \times (5.70 - 4.82)).

An alternative explanation for the positive coefficients on \(\text{Growth} \cdot \text{Ln}($Stdev) \cdot y\) and \(\text{Size} \cdot \text{Ln}($Stdev) \cdot y\) is that as a firm grows and increases in size, it becomes more diversified and \$Stdev is
more likely to capture the systematic risk. If the average \(PPS\) is less negative for firms with higher
systematic risk (compared to a similar level of firm idiosyncratic risk), then positive coefficients
for \(\text{Growth} \cdot \text{Ln}($Stdev) \cdot y\) and \(\text{Size} \cdot \text{Ln}($Stdev) \cdot y\) could capture the diversification effect. To
control for this effect, Table 3 replicates the main results in Table 2 with two additional control
variables. The first (\(\text{Herfindahl}\)) is a firm’s Herfindahl index (ranging between zero and one)
based on the firm’s sales revenues from its business segments during the previous year. This is
a standard measure of the degree of diversification in a firm’s operations, with higher values of
the index indicating lower degrees of diversification. We obtain qualitatively similar results when
we use the number of business segments of a firm to proxy for the firm’s diversification. The
second control variable (\(R2\)) is the systematic component of a firm’s return variation, calculated
as the \(R\)-squared from regressing a firm’s daily return on the market index and an industry index
during the previous year. As Table 3 shows, the coefficients for both \(\text{Growth} \cdot \text{Ln}($Stdev) \cdot y\)
and \(\text{Size} \cdot \text{Ln}($Stdev) \cdot y\) remain significantly positive (at less than the 5% level) even after including
these control variables.33

In summary, the empirical results are generally consistent with our model’s predictions. To the

33Since we use the \(\ln($Stdev)\), \(R2\) is included in the form of \(R2 \cdot y\), as \(\ln(R2 \cdot $Stdev)\) is linearly additive in \(R2\)
and \$Stdev.
best of our knowledge, no prior study has generated (in one setting) the dual predictions concerning the relations between growth/firm size and the cross-sectional variation in PPS sorted both by CEO age (tenure) and by firm risk.

5 Conclusion

This paper revisits the relation between managers’ career concerns and the optimal pay-for-performance sensitivity in their explicit compensation contracts. Prior studies (e.g., Fama (1980), Holmstrom (1999), Gibbons and Murphy (1992)) suggest that the power of explicit contracts can be lowered when managers are subject to reputation concerns in the external labor market. We show that not all career concerns substitute for explicit contracts. In particular, in a dynamic, multitask agency model, we find that when a manager is concerned about both the market’s perception of his ability and the variation in such perception, the substitutive relation between implicit and explicit incentives may be weakened, or even reversed. Further, the optimal pay-performance sensitivity may increase in firms’ underlying risk measures. Using a large sample of CEO compensation data from ExecuComp, we document evidence that is consistent with our model predictions. Our findings suggest that the external labor market affects the internal labor market contract in more complicated ways than prior studies suggest. Our analysis also has implications for the debate on the relation between firm risk and pay-for-performance sensitivity.
Appendix

A1 Proof of Lemma 1:

Proof. Because the agent chooses \( a \) before \( s \) is realized, his choice of \( a \) does not affect the expected level of the compensation. \( E(y_1|s) = E(\pi_1 + e_1 + \varepsilon_1|s) = E(\pi_1|s) + e_1 = as + e_1 \), where the second equality follows because \( \varepsilon_1 \) and \( s \) are independent. Note that the effective risk that the agent will face under equation (5) is proportional to

\[
\beta_1^2 \text{Var}(y_1 - as) = \beta_1^2 \left( \text{Var}(y_1 - as) + (a - \bar{a})^2 \text{Var}(s) \right) = \beta_1^2 \left( (1 - a) \sigma_\theta^2 + \frac{(a - \bar{a})^2}{a} \sigma_\theta^2 + \sigma_\varepsilon^2 \right).
\]

The first equality holds because \( y_1 - as = y_1 - E(y_1|s) + e_1 \) is independent of \( s \). Consequently, the agent’s objective function (excluding the terms unrelated to \( e_1, a \) and \( \bar{a} \)) is given by:

\[
\max_{e_1, a, \bar{a}} \beta_1 e_1 - c(e_1) - g(a) - \frac{r}{2} \beta_1^2 \left( (1 - a) \sigma_\theta^2 + \frac{(a - \bar{a})^2}{a} \sigma_\theta^2 + \sigma_\varepsilon^2 \right). 
\]

Setting the first-order condition for \( \bar{a} \) to zero yields the agent’s optimal announced \( \bar{a}^* = a \). Q.E.D.

A2 Proof of Lemma 2:

Proof. First, solving the agent’s maximization problem in period two (the IC constraint) yields \( e_2^* = k\beta_2 \). Substituting this into the principal’s objective function and setting the first-order-condition w.r.t. \( \beta_2 \) to zero gives the optimal \( \beta_2^* = \frac{1}{1+\rho Var(\frac{1}{2}(s_1,y_1))} \). This is a standard result of linear contracts with normal-distributed signals (see, e.g., GM). Second, the expected second-period total surplus, \( TS_2 \), can be expressed as

\[
TS_2 = kE(\pi_2|s_1,y_1) + V_2, \\
\text{where } V_2 = ke_2 - c(e_2) - \frac{r}{2} \beta_2^2 \text{Var}(y_2|s_1,y_1). 
\]
The first term in (12) is the excess output attributable to the agent’s ability:

\[ kE(\pi_2|s, y_1) = kpE(\theta|s, y_1), \quad (14) \]

where \( E(\theta|s, y_1) \equiv \hat{\theta} = h_y (y_1 - \hat{e}_1) + h_s s + (1 - h_y - h_s)E_0(\theta) \)

with \( h_y = \frac{(1 - a)p_\epsilon}{p_\theta + (1 - a)p_\epsilon}, h_s = \frac{ap_\theta}{p_\theta + (1 - a)p_\epsilon} \).

(14) shows that the manager’s expected ability at the beginning of period two is a weighted sum of three signals: the first-period output \( y_1 \), the signal \( s \), and the prior information about \( \theta \) as summarized by \( E_0(\theta) \). Since \( E_0(\theta) \) is normalized to zero, the last term, \( (1 - h_y - h_s)E_0(\theta) \), drops out. \( \hat{e}_1 \) is the agent’s first-period effort choice perceived by the market, and is, in equilibrium, the actual effort that the agent supplies.

The second term in (12), \( V_2 \), is the contracting surplus from the relationship. Substituting \( \beta_2^* \) and \( e_2^* \) in (13) yields

\[ V_2 = \frac{k^2}{2[1 + rVar(y_2^0|s, y_1)]}, \quad \text{where } Var(y_2^0|s, y_1) = \rho^2Var(\theta|y_1, s) + Var(\varepsilon_2^0). \]

Simple algebra shows that \( \frac{\partial V_2}{\partial a} > 0 \) as follows:

\[ \frac{\partial V_2}{\partial a} = \frac{rk^2\rho^2}{2[1 + r(\rho^2Var(\theta|y_1, s) + Var(\varepsilon_2^0))]^2} - \frac{\partial Var(\theta|y_1, s)}{\partial a} > 0, \tag{15} \]

where \( \varepsilon_2' = \sqrt{1 - \rho^2}\varepsilon + \varepsilon_2 \), and \( Var(\theta|y_1, s) = \frac{(1 - a)}{p_\theta + (1 - a)p_\epsilon} \).

It is straightforward to see that \( \frac{\partial V_2}{\partial a} \) is increasing in \( k \).

To obtain the final expression of (6), notice that the optimal \( \alpha_2^* \) can be obtained by substituting \( TS_2 \) into the agent’s participation constraint, using the fact that the constraint holds with equality at \( \alpha_2^* \). Q.E.D. □

A3 Derivation of equation (7)

Note that

\[ E_0[y_2 - E(y_2|s, y_1)] = k\left[ \rho E_0(\theta) + e_2^* - \rho E_0(\hat{\theta}) \right] - e_2^* = -\rho kE_0(\hat{\theta}), \]

where \( e_2^* \) is the optimal production effort that the agent expects to expend in the second period, and \( \hat{\theta} \) is the principal’s perception of the agent’s ability after period one (given by (14)). Therefore,
taking expectation of $w_2$ as expressed in (6) yields

$$E_0(w_2) = ho k (b - \beta_2) E_0 \left( \hat{\theta} \right) + \frac{r}{2} Var(w_2|s,y_1) + bV_2 + c(e^*_2).$$

Note that $E_0 \left( \hat{\theta} \right) = h_y (e^*_1 - \hat{e}_1)$ is the agent’s expectation (formed at the beginning of period one) of his perceived ability in the second period, where $e^*_1$ is the optimal production effort that the agent expects to expend in the first period, and $\hat{e}_1$ is the market’s assessment of $e_1$ (the two are equal in equilibrium).

Next, we use the identity $Var_0(w_2) = Var(w_2|s,y_1) + Var_0[E(w_2|s,y_1)]$ (that is, the total variance of a random variable equals the sum of the conditional variance and variance of the conditional mean). Substitution yields (7) as the agent’s objective function.

A4 Proof of Lemma 3:

Proof. It suffices to show that the right hand side of equation (9) is increasing in $\beta_1$. This is straightforward because

$$\frac{\partial \text{RHS of (9)}}{\partial \beta_1} = \frac{-r \partial^2 Var_0(w_1)}{2 \partial a \partial \beta_1} = \frac{r \beta_1 \sigma^2_\theta}{\sigma^2_\theta} > 0.$$

A5 Proof of $\frac{\partial Var_0[E(w_2|s,y_1)]}{\partial a} > 0$:

Proof. Because

$$Var_0[E(w_2|s,y_1)] = b^2 \rho^2 k^2 Var_0 \left( \hat{\theta} \right),$$

it suffices to show that $\frac{\partial Var_0(\hat{\theta})}{\partial a} > 0$. This is true because

$$Var_0 \left( \hat{\theta} \right) \equiv Var_0[E(\theta|s,y_1)] = \frac{1}{p_\theta} \frac{(1-a)p_\varepsilon + ap_\theta}{p_\theta + (1-a)p_\varepsilon} = (h_y + h_s) \sigma^2_\theta.$$

where $h_y$ and $h_s$ are defined in (14), and

$$\frac{\partial Var_0(\hat{\theta})}{\partial a} = \frac{p_\theta}{[p_\theta + (1-a)p_\varepsilon]^2} > 0.$$

This completes the proof. □
A6 Proof of Proposition 1:

Proof. At the beginning of period one, the principal chooses $\beta_1$ to maximize the total expected surplus, subject to the agent’s first-order conditions ((8) and (9)), excluding terms unrelated to $\beta_1, e_1$ and $a$:

$$
\max_{\beta_1, e_1, a} f = \left\{ e_1 - c(e_1) - g(a) - \frac{r}{2} \beta_1^2 \left[ (1 - a) \sigma_\theta^2 + \sigma_\varepsilon^2 \right] \right\} + \left[ V_2 - \frac{r}{2} b^2 \rho^2 k^2 \text{Var}_0 \left( \frac{\hat{\theta}}{\theta} \right) \right].
$$

(18)

The first-order condition with respect to $\beta_1$, taking into account the agent’s optimal responses, is given by

$$
\frac{df}{d\beta_1} = \frac{de_1}{d\beta_1} - c'(e_1) \frac{de_1}{d\beta_1} - g'(a) \frac{da}{d\beta_1} - r \beta_1 \left[ (1 - a) \sigma_\theta^2 + \sigma_\varepsilon^2 \right] + \frac{r}{2} \beta_1^2 \sigma_\theta^2 \frac{da}{d\beta_1} + \frac{\partial}{\partial a} \left[ \frac{\partial V_2}{\partial a} - \frac{r}{2} b^2 \rho^2 k^2 \text{Var}_0 \left( \frac{\hat{\theta}}{\theta} \right) \right] \frac{da}{d\beta_1}.
$$

Substituting in the agent’s first-order conditions yields:

$$
\frac{df}{d\beta_1} = \left( 1 - \beta_1 \right) \frac{de_1}{d\beta_1} - r \beta_1 \left[ (1 - a) \sigma_\theta^2 + \sigma_\varepsilon^2 \right] - \rho k (b - \beta_2) h_\theta \frac{de_1}{d\beta_1} + \frac{r}{2} b^2 \rho^2 k^2 \text{Var}_0 \left( \frac{\hat{\theta}}{\theta} \right) \frac{da}{d\beta_1}.
$$

(19)

Combining the last two terms yields:

$$
\frac{df}{d\beta_1} = \left( 1 - \beta_1 \right) \frac{de_1}{d\beta_1} - r \beta_1 \left[ (1 - a) \sigma_\theta^2 + \sigma_\varepsilon^2 \right] - \rho k (b - \beta_2) \frac{de_1}{d\beta_1} + \frac{r}{2} b^2 \rho^2 k^2 \text{Var}_0 \left( \frac{\hat{\theta}}{\theta} \right) \frac{da}{d\beta_1}.
$$

(20)

Using the envelope theorem, we know that $\beta_1$ is increasing (decreasing) in $\rho$ if $\frac{df}{d\beta_1} / \partial \rho > (<) 0$. If we can show that (i) as $k \rightarrow \infty$, $\partial \frac{df}{d\beta_1} / \partial \rho > 0$; (ii) when $k \rightarrow 0$, $\partial \frac{df}{d\beta_1} / \partial \rho < 0$, then by continuity, as $k$ gets larger, the negative relation between $\beta_1$ and $\rho$ is weakened and eventually turns positive.
Notice that $\frac{\partial A}{\partial \rho}$ is well-defined and bounded. The absolute value of $\frac{\partial B}{\partial \rho}$ increases with $k$ at the order of $k$. Finally, we look at the (C) term. Let $\sigma_\theta^2 = Var(\pi_2|y_1,s)$. Given that $V_2 = \frac{k^2}{2[1+rVar(\frac{\pi_2}{k}|y_1,s)]} = \frac{k^2}{2}\beta_2$, we have

$$
\frac{\partial C}{\partial \rho} = (1-b)\frac{\partial V_2}{\partial a} \frac{da}{\partial \beta_2} = (1-b) \frac{k^2}{2} \left[ \frac{\partial \beta_2}{\partial a} \frac{da}{\partial \beta_2} \right] \\
= (1-b) \frac{k^2}{2} \left( -\frac{\partial C}{\partial a} \right) \frac{\partial \left[ \beta_2^2 \left( \rho^2 \frac{da}{\partial \beta_2} \right) \right]}{\partial \rho}.
$$

Notice the following results: (i) $\frac{\partial \sigma_\theta^2}{\partial a} = \frac{-\rho}{|\rho_0 + (1-a)p_e|} < 0$. (ii) $\frac{\partial \beta_2^2}{\partial \rho} < 0$, as

$$
\frac{\partial \beta_2}{\partial \rho} = \frac{-r}{\left[ 1 + rVar \left( \frac{\pi_2}{k} | y_1,s \right) \right]^2} \frac{\partial Var \left( \frac{\pi_2}{k} | y_1,s \right)}{\partial \rho} = \frac{2r\rho\beta_2}{\left( \sigma_\theta^2 - \tilde{\sigma}_\theta^2 \right)} > 0.
$$

And (iii) $\frac{da}{d\beta_2}|_{a=a^*} > 0$. This follows immediately by taking implicit differentiation over (9) and noticing that the optimal $\tilde{a} = a$ and in equilibrium $e_1^* = \tilde{e}_1$, i.e., $\frac{da}{d\beta_2}|_{a=a^*} = \frac{\partial FOC}{\partial a} = \frac{\partial Var(\pi_2)}{\partial \rho} > 0$, where SOC stands for the second-order condition with respect to $a$. The SOC is non-positive assuming that an interior solution for $a$ exists.

Thus, as long as we can show $\frac{\partial}{\partial \rho} \left( \frac{\rho^2}{a^2} \right) > 0$, we know that $\frac{\partial C}{\partial \rho}$ is increasing at the order of $k^2$. Since $\rho^2 \frac{da}{d\beta_2} = \frac{r\beta_2\sigma_\theta^2}{(-SOC/\rho^2)}$, we have $\frac{\partial}{\partial \rho} \left( \frac{\rho^2}{a^2} \right) = \frac{r\beta_2\sigma_\theta^2}{(-SOC/\rho^2)}$. To see that $\left( SOC/\rho^2 \right)$ indeed increases with $\rho$, notice that

$$
\frac{\partial}{\partial \rho} \left( SOC/\rho^2 \right) = \frac{\partial}{\partial \rho} \left[ -g''(a)/\rho^2 - \frac{\partial}{\partial a} \left( d\rho \right) \frac{\partial}{\partial \beta_2} \frac{da}{d\beta_2} \right] / \rho^2 + b^2 \frac{\partial}{\partial \rho} \frac{V_2}{\rho^2}.
$$

$$
= \frac{\partial}{\partial \rho} \left[ -g''(a)/\rho^2 \right] b - \frac{r}{2} \frac{\partial}{\partial \rho} \left[ k^2 \frac{\partial}{\partial \rho} \frac{\partial}{\partial \beta_2} \frac{da}{d\beta_2} \right] + b^2 \frac{k^2}{2} \frac{\partial}{\partial \rho} \left[ \frac{\partial^2}{\partial \beta_2^2} \frac{da}{d\beta_2} \right] / \rho^2.
$$

The first term is obviously increasing in $\rho$. Based on $\beta_2^* = \frac{1}{1+rVar(\frac{\pi_2}{k}|y_1,s)}$, we have

$$
\frac{\partial^2 \beta_2}{\partial a^2} / \rho^2 = r \left[ - \left( 2\beta_2 \frac{\partial \beta_2}{\partial a} \frac{\partial \sigma_\theta^2}{\partial a} + \beta_2^2 \frac{\partial^2 \sigma_\theta^2}{\partial a^2} \right) \right] > 0,
$$

34
due to the fact that $\frac{\partial^2 \beta}{\partial a^2} > 0$, $\frac{\partial^2 \sigma^2}{\partial a^2} < 0$, and $\frac{\partial^2 \sigma^2}{\partial a^2} < 0$. Further, the magnitude of each term is increasing in $\rho$. Hence $\frac{\partial^2}{\partial a^2} \left[ \frac{\partial v_k}{\partial a} \right]$ is increasing with $\rho$ (at the order of $k^2$).

In summary, we have shown that $\frac{\partial C}{\partial \rho}$ is positive and increases at the order of $k^2$. Combining the (A), (B) and (C) terms, we see that the (C) term will dominate when $k$ is sufficiently large. Therefore, $\frac{\partial f}{\partial \sigma^2} > 0$ when $k$ is sufficiently large.

On the other hand, when $k \to 0$, $\frac{\partial B}{\partial \rho}$ dominates $\frac{\partial C}{\partial \rho}$ because the former is of order $O(k)$, and the latter is of order $O(k^2)$. Therefore, $\frac{\partial f}{\partial \sigma^2} < 0$ when $k$ is sufficiently small. This completes the proof.

\section{Proof of Proposition 2:}

\textbf{Proof.} The proof is analogous to that of Proposition 1. Starting with equation (20), we want to show that as $\sigma^2_\theta \to \infty$ and $k \to \infty$, it is the case that

$$\frac{\partial f}{\partial \sigma^2_1} > 0.$$  

>From before, we know that $\frac{\partial A}{\partial \sigma^2_\theta}$ is well-defined and bounded, and that the absolute value of $\frac{\partial B}{\partial \rho}$ increases with $k$ at the order of $k$. Finally, we look at the (C) term:

$$\frac{\partial C}{\partial \sigma^2_\theta} = (1 - b) \frac{\partial}{\partial \sigma^2_\theta} \frac{\partial v_2}{\partial \sigma^2 \sigma_\theta} \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} = (1 - b) k^2 \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} \left( \ast \right).$$

We want to show that $(\ast)$ is positive. Note that $\frac{\partial}{\partial \sigma^2_\theta} \left( \frac{1}{z(1 + r[\sigma^2_\theta - \rho^2 \text{Var}_0(\hat{\theta}) + \sigma^2])} \right) \to 0$ when $\sigma^2_\theta \to \infty$,

therefore sign $\left( \frac{\partial}{\partial \sigma^2_\theta} \left( \frac{1}{z(1 + r[\sigma^2_\theta - \rho^2 \text{Var}_0(\hat{\theta}) + \sigma^2])} \right) \right) = \text{sign} \left( \frac{\partial}{\partial \sigma^2_\theta} \frac{\partial d \text{Var}_0(\hat{\theta})}{d \beta_1} \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} \right)$. Next,

$$\frac{\partial d \text{Var}_0(\hat{\theta})}{d \beta_1} \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} \frac{d \text{Var}_0(\hat{\theta})}{d \beta_1} = \frac{\partial}{\partial \sigma^2_\theta} \frac{\partial}{\partial \sigma^2_\theta} \frac{d a}{d \beta_1} \frac{d a}{d \beta_1} = \frac{\partial}{\partial \sigma^2_\theta} \frac{d a}{d \beta_1} \frac{d a}{d \beta_1}.$$

35
Note that \( \frac{p_\theta}{[p_\theta + (1-a)p_\varepsilon]} \) is increasing in \( \sigma_\theta^2 \) when \( \sigma_\varepsilon^2 \) is large, and by Lemma 3, \( \frac{da}{d\sigma_1} \) is positive. If we can show that \( \frac{da}{d\sigma_1} \) is increasing in \( \sigma_\theta^2 \), then \( \frac{\partial \text{Var}_0(\bar{\theta})}{\partial \sigma_\theta^2} > 0 \) when \( \sigma_\theta^2 \) is large, and hence we know (21) increases at the order of \( k^2 \) when \( \sigma_\theta^2 \) is large.

Following similar steps as in the proof for Proposition One, we have

\[
\frac{\partial}{\partial \sigma_\theta^2} \left[ \frac{d_\theta}{d\sigma_1} \right] = \frac{-\beta_1}{\langle -\text{SOC} \rangle} \left( \sigma_\theta^2 \frac{\partial}{\partial \sigma_\theta^2} \langle -\text{SOC} \rangle \right) - \frac{\beta_1}{\langle -\text{SOC} \rangle^2} \langle -\text{SOC} \rangle^2.
\]

Notice that

\[
\frac{\partial}{\partial \sigma_\theta^2} \left( \langle -\text{SOC} \rangle^2 \right) = \frac{\partial}{\partial \sigma_\theta^2} \left( \langle -\text{SOC} \rangle \right)^2 = 2 \langle -\text{SOC} \rangle \left( \frac{\partial}{\partial \sigma_\theta^2} \langle -\text{SOC} \rangle \right) + \left( \frac{\partial^2}{\partial \sigma_\theta^2} \langle -\text{SOC} \rangle \right)^2.
\]

We know that both \( \frac{\partial^2}{\partial \sigma_\theta^2} \langle -\text{SOC} \rangle \) and \( \frac{\partial}{\partial \sigma_\theta^2} \langle -\text{SOC} \rangle^2 \) are increasing in \( \sigma_\theta^2 \) when \( \sigma_\theta^2 \) is large (detailed derivations shown at the end of this section) and

\[
\left( \frac{\partial}{\partial \sigma_\theta^2} \langle -\text{SOC} \rangle \right)^2 = \beta_2 - \beta_2 r \sigma_\theta^2 \frac{\partial^2}{\partial \sigma_\theta^2} \langle -\text{SOC} \rangle
\]

is also increasing in \( \sigma_\theta^2 \) (since \( \beta_2 \) is decreasing in \( \sigma_\theta^2 \)). Therefore, \( \frac{\partial}{\partial \sigma_\theta^2} \langle -\text{SOC} \rangle \) is large, and hence we know (21) increases at the order of \( k^2 \) when \( \sigma_\theta^2 \) is large.

Combining the (A), (B) and (C) terms, we see that the (C) term will dominate when \( k \) and \( \sigma_\theta^2 \) are sufficiently large, and therefore \( \frac{da}{d\sigma_1} \frac{\partial}{\partial \sigma_\theta^2} \sigma_\theta^2 > 0 \). This completes the proof.

Some detailed derivations used in the proof:

\[
\frac{\partial^2}{\partial \sigma_\theta^2} \frac{\partial}{\partial \sigma_\varepsilon^2} \langle -\text{SOC} \rangle = \frac{\partial^2}{\partial \sigma_\theta^2} \langle -\text{SOC} \rangle^2
\]

\[
\frac{\partial}{\partial \sigma_\theta^2} \langle -\text{SOC} \rangle = \frac{\partial}{\partial \sigma_\theta^2} \left( \frac{\partial^2}{\partial \sigma_\theta^2} \langle -\text{SOC} \rangle \right) + \frac{\partial}{\partial \sigma_\theta^2} \left( \frac{\partial^2}{\partial \sigma_\varepsilon^2} \langle -\text{SOC} \rangle \right)
\]

A8 Proof of the sufficiency of the two conditions regarding the signal for \( a \):
**Proof.** Let \( f(\omega) \) be the distribution for the signal for the agent’s \( a \) effort \((\omega = a + \varepsilon_\omega, \varepsilon_\omega \sim N(0, 1/p_\omega))\). The first condition, that \( \omega \) does not directly depend on the realized value of \( s \) ensures that \( f(\omega) = f(\omega|s) \). Subramanyam (1996) shows that under this condition, the principal will form a belief about \( \theta \) similar to (14) with \( a \) replaced with \( \omega \), that is,

\[
E(\theta|s, y_1) \equiv \hat{\theta} = h'_y (y_1 - \hat{c}_1) + h'_s s + (1 - h'_y - h'_s) E_0(\theta)
\]

\[
= h'_y (y_1 - \hat{c}_1) + h'_s s
\]

with

\[
h'_y = \frac{(1 - \omega)p_\varepsilon}{p_\theta + (1 - \omega)p_\varepsilon},
\]

\[
h'_s = \frac{\omega p_\theta}{p_\theta + (1 - \omega)p_\varepsilon}.
\]

Therefore, higher \( a \), in expectation, leads to higher \( \text{Var}_0(\hat{\theta}) \) via \( \omega \). Further, the second condition is that the agent does not control the precision of \( \omega (p_\omega) \). It requires that \( p_\omega \) is exogenous to the agent, and therefore the additional variation in \( \text{Var}_0(\hat{\theta}) \) due to noise \( \varepsilon_\omega \) is exogenous to the agent and does not enter his optimization. As a result, all our predictions in Propositions 1 and 2 go through. □
References


Core, John, and Jun Qian, 2001, Option-like contracts for project selection and production, Working paper, University of Pennsylvania and Boston College.


Hirshleifer, David, Tarun Chordia, and Seongyeon Lim, 2001, Firm and managerial incentives to manipulate the timing of project resolution, Working paper, Ohio State University and Emory University.


