Abstract

We introduce human capital accumulation in a life-cycle model of career concerns and analyze its implications for the optimal provision of incentives. We first show that human capital accumulation increases implicit incentives, but that these incentives disappear over time even when uncertainty about talent is exogenously replenished. Second, we show that it can be optimal for explicit incentives to complement rather than substitute implicit incentives. Overall, our results help explain the conflicting evidence on the sensitivity of pay to performance for managers at different stages of their working life.

1 Introduction

Fama (1980) famously argued that a worker’s concern for his reputation in the labor market can alone remove moral hazard problems, making the use of explicit incentive contracts unnecessary. More precisely, Fama hypothesized that if individual productivity is uncertain, then an individual’s concern for how his performance affects the labor market’s assessment of his productivity, and thus his future wages, may be sufficient to provide appropriate incentives for effort. Holmström (1982,1999) (Holmström hereafter) formalized Fama’s intuition
and proved that implicit ‘career concern’ incentives can indeed alleviate moral hazard, but that these incentives disappear as the market’s assessment of the worker’s ability becomes more precise. Thus, unless uncertainty about talent is constantly replenished, the effect of implicit incentives on performance is only temporary and hence explicit incentives are needed. Then, a natural question to ask is what is the optimal provision of explicit incentives in the presence of implicit incentives. Gibbons and Murphy (1992) (Gibbons and Murphy hereafter) showed that if uncertainty about talent is decreasing over time, then explicit incentives should be used as a substitute for career concerns. In particular, explicit incentives should be strongest for individuals at the end of their working life.

In this paper we analyze how the provision of implicit incentives and the relationship between implicit and explicit incentives are affected by the possibility for workers to acquire human capital when employed. Our starting point is a finite–horizon version of Holmström where a worker’s ability improves over time through a learning–by–doing component of effort: the more effort an individual exerts, the greater the amount of additional skills he acquires. We consider a finite–horizon framework in order to focus on the life–cycle profile of incentives (implicit and explicit).

The first result we obtain is that the presence of human capital accumulation increases implicit incentives. However, these incentives disappear with a worker’s age even in the presence of exogenous replacements of uncertainty. This result, probably not that surprising, is due to the fact that as a worker gets older, and so his future career gets shorter, the benefit he obtains from influencing the market’s assessment about his ability diminishes. We refer to this effect on implicit incentives as the tenure effect, in order to distinguish it from the effect, emphasized by Holmström, of the precision about a worker’s ability on his career concerns.

We then consider, as in Gibbons and Murphy, the environment where implicit incentives can be supplemented by explicit incentives provided by performance–based contracts that are linear in output. We show that unlike in Gibbons and Murphy, optimal total incentives can be decreasing (as opposed to increasing) over time and that it can be optimal to use implicit and explicit incentives as complements rather than substitutes. In particular, it can be optimal for explicit incentives to be the greatest earlier in a worker’s career, when implicit incentives are also the highest.
Overall, our results help explain the conflicting evidence on the importance of explicit performance–based incentives for managers at different stages of their working life. For instance, Jensen and Murphy (1990) find small sensitivity of pay to performance for CEO’s. Instead, Barro and Barro (1990) report a stronger pay–for–performance sensitivity for younger than for older executives: they document that the sensitivity of changes in pay to performance decreases as CEO experience increases. Analogously, using longitudinal data on managers from a single firm, Khan and Sherer (1990) find that bonuses paid to managers who are assigned to high–level positions, work at corporate headquarters, and have low seniority are more sensitive to performance than bonuses for managers without those three characteristics. Gibbons and Murphy (1992) and Garvey and Milbourn (2003) report a stronger positive relationship between pay and performance for older than for younger executives. Gompers and Lerner (1999) obtain similar results in venture capital limited partnership agreements: the compensation of new and smaller funds displays less sensitivity to performance than that of other funds. Our results suggest that the strength and time profile of the relationship between performance and explicit incentives depend on the importance of the process of human capital accumulation on the job.

The rest of the paper is organized as follows. In Section 2 we analyze the benchmark career concerns model. In Section 3 we study the case where the firms in the market can offer short–term linear contracts. In section 4 we discuss in more detail the relationship between our work and the literature (theoretical and empirical). We conclude in Section 6. The Appendix contains omitted proofs and details.

2 Benchmark Model

In this section we describe the benchmark environment, characterize its unique equilibrium, and study the lifetime profile of effort.
2.1 Basic Setup

We consider a worker in a competitive labor market. Time is discrete and begins in period \( t = 1 \). The worker lives for \( T \geq 2 \) periods and has discount factor \( \delta \in (0, 1] \). The worker’s output in period \( t \) is

\[ y_t = a_t + \theta_t + k_t + \varepsilon_t, \]

where \( a_t \) is his private choice of effort (labor input), \( \theta_t \) is his ability, \( k_t \) is his (general) human capital, and \( \varepsilon_t \) is a noise term. The noise terms are independent and normally distributed with mean zero and precision \( h_\varepsilon \). The worker’s human capital in period \( t \) depends on his past labor inputs. We assume that for all \( t \geq 2 \),

\[ k_t = k_t(a_1, \ldots, a_{t-1}) = k_1 + k(a_1) + \cdots + k(a_{t-1}), \]

where \( k_1 \) is the worker’s initial human capital and the function \( k \) is three times differentiable and concave, with \( k(0) = 0 \) and \( k''' \leq 0 \). The worker’s ability is unknown to both him and the market and evolves according to

\[ \theta_{t+1} = \theta_t + \eta_t, \]

where the shock terms \( \eta_t \) are independent and normally distributed with mean zero and precision \( h_\eta \). We discuss the assumption of shocks to ability later in this section. The worker and the market have a common prior belief about \( \theta_1 \) that is normally distributed with mean \( m_1 \) and precision \( h_1 \). We refer to the market’s belief about \( \theta_t \) as the worker’s reputation in period \( t \).

Output–contingent contracts are not possible, and so the worker’s pay in a period cannot depend on his output in that period. Let \( w_t \) be the worker’s wage in period \( t \). The worker’s payoff from a sequence \( \{a_t\}_{t=1}^T \) of efforts and a sequence \( \{w_t\}_{t=1}^T \) of wage payments is

\[ -\exp \left( -r \left\{ \sum_{t=1}^T \delta^{t-1} [w_t - g(a_t)] \right\} \right), \]

where \( r > 0 \) is the worker’s coefficient of absolute risk aversion and \( g(a) \) is his disutility from effort \( a \). We assume that \( g \) is three times differentiable, strictly increasing, and strictly

---

\(^{1}\)We can easily adapt our analysis to the case where human capital depreciates, as in \( k_t(a_1, \ldots, a_{t-1}) = \lambda^{t-1}k_1 + \sum_{s=1}^{t-1} \lambda^{t-s-1}k(a_s), \) where \( \lambda \in (0, 1) \) is the depreciation rate of human capital.
convex, with \( g'' \geq 0, \ g'(0) = 0, \) and \( g'(\infty) = \infty. \) This specification of preferences, which is the same as in Gibbons and Murphy, allows us to focus on the tradeoff between risk and incentives when we study the optimal provision of explicit incentives.

Also as in Gibbons and Murphy, we allow effort to be negative with \( g'(-\infty) = -\infty \) and interpret negative effort as the worker either stealing or hiding output. We assume that such activities destroy human capital: \(-k(a)\) is the amount of human capital the worker destroys when \( a < 0. \) Since \( g \) is strictly convex and \( g'(0) = 0, \) stealing or hiding output becomes increasingly more difficult on the margin. Notice that the assumptions that \( g''' \geq 0 \) and \( g \) is strictly convex are incompatible with \( g''(0) = 0 \) when allow for negative effort. Hence, we assume that \( g''(0) > 0. \) As we show in the next subsection, the worker never exerts negative effort in the benchmark career concerns case. In Section 3, we discuss conditions under which the worker’s effort is always non–negative in the presence of short–term linear contracts.

Let \( Y_t = \mathbb{R}_t^{t-1}, \) with typical element \( y^t, \) be the set of period–t output histories, \( A_t = \mathbb{R}_t^{t-1}, \) with typical element \( a^t, \) be the set of period–t labor input histories, and \( Z_t = Y_t \times A_t, \) with typical element \( z^t = (a^t, y^t), \) be the set of period–t (worker) histories. A strategy for the worker is a sequence \( \sigma = \{\sigma_t\}_{t=1}^T, \) with \( \sigma_t : Z_t \to \Delta(\mathbb{R}_+), \) such that \( \sigma_t \) is the worker’s (mixed) choice of effort in period \( t \) as a function of his past action choices and outputs.\(^2\) We say the strategy \( \sigma = \{\sigma_t\} \) is uncontingent if \( \sigma_t \) is constant in \( Z_t \) for \( t = 1, \ldots, T. \)

A strategy for the market is a sequence \( \omega = \{\omega_t\}_{t=1}^T, \) with \( \omega_t : Y_t \to \mathbb{R}, \) such that \( \omega_t(y^t) \) is the wage the worker receives in period \( t \) if his output history is \( y^t. \) A strategy \( \sigma \) for the worker is sequentially rational given a strategy \( \omega \) for the market if it maximizes the worker’s lifetime payoff after every history. Since the market is competitive, and so firms make zero (expected) profits, the worker’s wage in period \( t \) is his expected output in that period. In other words, if the market expects the worker to follow the strategy \( \sigma \) and the worker’s period–t output history is \( y^t, \) then \( w_t = \mathbb{E}[y_t|\sigma, y^t]. \)

**Definition 1.** An equilibrium is a pair \((\sigma^*, \omega^*)\) such that \( \sigma^* \) is sequentially rational given \( \omega^* \) and \( w^*_t(y^t) = \mathbb{E}[y_t|\sigma^*, y^t] \) for all \( t \geq 1 \) and \( y^t \in Y^t. \) We say the equilibrium is uncontingent if \( \sigma^* \) is uncontingent.\(^2\)

\(^2\)In principle, the worker could also condition his behavior on his past wages. Since the wage determination process is non–strategic, there is no reason for him to do so.
We restrict attention to equilibria where the worker follows a pure strategy. In the next subsection we show that there exists a unique equilibrium and provide a complete characterization of this equilibrium.

2.2 Career Concerns with Human Capital Accumulation

Suppose the worker follows the strategy $\sigma = \{\sigma_t\}$ and let $y_t$ and $k_t$ be his period–$t$ output history and human capital, respectively. Suppose also that the worker’s reputation in period $t$ is normally distributed with mean $m_t$ and precision $h_t$, and let

$$\mu_t = \frac{h_t}{h_t + h_x}.$$  \hfill (2)

A standard argument shows that if the worker’s output in period $t$ is $y_t$, then his reputation in period $t + 1$ is normally distributed with mean $m_{t+1}$ and precision $h_{t+1}$, where

$$m_{t+1} = m_{t+1}(y^t, y_t) = \mu_t m_t + (1 - \mu_t)(y_t - \sigma_t(y^t) - k_t)$$  \hfill (3)

and

$$h_{t+1} = \frac{(h_t + h_x) h_x}{h_t + h_x + h_\eta}.$$  \hfill (4)

Since the prior belief about $\theta_1$ is normally distributed, we then have that regardless of the strategy the worker follows, his reputation at any point in time is normally distributed. This is also true of the worker’s posterior belief about his ability (which coincides with his reputation on the path of play). Equation (4) implies that the evolution of the precision $h_t$ is deterministic and independent of the worker’s behavior.

In order to understand the impact of learning by doing on implicit incentives, consider the case where the market expects the worker to follow an uncontingent strategy $\tilde{\sigma}$ and let $\tilde{a}_t$ be the worker’s conjectured choice of effort in period $t$. By the zero–profit condition, the worker’s wage in period $s \geq 2$ is

$$w_s(y_1, \ldots, y_{s-1}) = m_s(y_1, \ldots, y_{s-1}) + \tilde{a}_s + \sum_{q=1}^{s-1} k(\tilde{a}_q);$$

besides the market’s conjecture about the worker’s effort in periods 1 to $s$, the worker’s wage in period $s$ also depends on his output in periods 1 to $s - 1$. A straightforward argument
(see the Appendix for a proof) shows that for all s and t with \( s \geq t + 1 \),

\[
m_s(y_1, \ldots, y_{s-1}) = m_s(y^t, y_t, \ldots, y_{s-1})
= m_t(y^t) \prod_{\tau=t}^{s-1} \mu_\tau + \sum_{q=t}^{s-1} (1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_\tau \left[ y_q - \hat{a}_q - \sum_{r=1}^{q-1} k(\hat{a}_r) \right],
\]

where we adopt the convention that \( \prod_{\tau=t}^{s-1} \mu_\tau = 1 \) when \( t = s \).

Note that the worker’s optimal choice of effort in period \( T \) is zero no matter the market’s conjecture about his behavior. Consider now the worker’s choice of effort \( a \) in period \( t \leq T-1 \) when his history is \( z^t \) and he behaves according to the uncontingent strategy \( \hat{\sigma} \) from period \( t + 1 \) on. For this, notice that if \( \mathbb{E}[\theta_q|z^t] \) denotes the mean of the worker’s posterior belief about his ability in period \( q \geq t \) given \( z^t \), then

\[
\mathbb{E}[y_q - \hat{a}_q - \sum_{r=1}^{q-1} k(\hat{a}_r)|a, z^t] = \begin{cases} 
\mathbb{E}[\theta_t|z^t] + a - \hat{a}_t & \text{if } q = t \\
\mathbb{E}[\theta_q|z^t] + k(a) - k(\hat{a}_q) & \text{if } q \geq t + 1
\end{cases}
\]

Therefore, for all \( s \geq t + 1 \),

\[
\frac{\partial \mathbb{E}[w_s|a, z^t]}{\partial a} = (1 - \mu_t) \prod_{\tau=t+1}^{s-1} \mu_\tau + \sum_{q=t+1}^{s-1} (1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_\tau k'(a).
\]

Equation (6) shows the impact of the worker’s choice of effort in period \( t \) on his (expected) wage payment in all subsequent periods. The worker’s effort in period \( t \) affects his wage in two ways. First, by affecting his output in period \( t \), the worker’s effort directly influences his reputation, and thus his wage, in all subsequent periods. This is the standard career concerns effect identified by Holmström and it corresponds to the term \( (1 - \mu_t) \prod_{\tau=t+1}^{s-1} \mu_\tau \) in (6). The second effect is due to the accumulation of human capital. Since the worker’s choice of effort in period \( t \) affects his human capital from period \( t + 1 \) on, it has an impact on his output from period \( t + 1 \) on. Thus, the worker’s choice of effort in period \( t \) has an additional impact on his reputation from period \( t + 2 \) on. This second effect, which corresponds to the term \( \sum_{q=t+1}^{s-1} (1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_\tau k'(a) \) in (6) and is only present when \( T \geq 3 \), increases with the marginal return of effort on the accumulation of human capital.

We know that if the random variable \( x \) is normally distributed with mean \( m \) and variance \( \sigma^2 \), then \( \mathbb{E}[\exp(-rx)] = \exp(-rm + r^2\sigma^2/2) \). It is easy to see that this implies that the
worker’s optimal choice of effort in period \( t \) is the unique solution \( a^*_t \) to
\[
\sum_{s=t+1}^{T} \delta^{s-t} \frac{\partial \mathbb{E}[w_s | a, z^t]}{\partial a} = \sum_{s=t+1}^{T} \delta^{s-t} \left[ (1 - \mu_t) \prod_{\tau=t+1}^{s-1} \mu_\tau + \sum_{q=t+1}^{s-1} (1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_\tau k'(a) \right] = g'(a); \quad (7)
\]
uniqueness follows from the concavity of \( k \) together with the strict convexity of \( g \). Since the left–hand side of (7) is positive, \( a^*_t \) is positive as well. Note that (7) also characterizes the worker’s optimal choice of effort in \( t = T \), in which case it reduces to \( g'(a) = 0 \), with unique solution \( a^*_T = 0 \). Since \( a^*_t \) is independent of \( z^t \) and the market’s conjecture \( \hat{\sigma} \), it is immediate to see that the pair \((\sigma^*, \omega^*)\), where \( \sigma^* \) is such that \( \sigma_t^*(z^t) \equiv a_t^* \) and \( \omega^* \) is such that \( \omega_t^*(y^t) = \mathbb{E}[y_t | \sigma^*, y^t] \), is the unique uncontingent equilibrium. Notice that \((\sigma^*, \omega^*)\) does not depend on the worker’s initial human capital \( k_1 \). It turns out that \((\sigma^*, \omega^*)\) is the unique equilibrium; see the Appendix for a proof. We have thus established the following result.

**Proposition 1.** There exists a unique equilibrium, which is uncontingent. The worker’s choice of effort in period \( t \geq 1 \) is the unique solution \( a^*_t \) to (7), which is non–negative.

Notice that when \( k' \equiv 0 \), and so there is no accumulation of human capital, Proposition 1 reduces to a finite–horizon version of Holmström with preferences given by (1). Since the right–hand side of (7) increases with \( k'(a) \), in every period \( t \leq T - 1 \) the equilibrium choice of effort is greater when there exists accumulation of human capital than when there exists no accumulation. In the next subsection we discuss the implications of the equilibrium condition (7) for the lifetime profile of effort and the efficiency of the worker’s behavior.

### 2.3 Discussion

We first discuss the equilibrium lifetime profile of effort. We know from Holmström that the sequence \( \{\mu_t\} \) given by (2) is such that
\[
\mu_{t+1} = \frac{1}{2 + \xi - \mu_t}, \quad (8)
\]
where \( \xi = h_\varepsilon / h_\eta \). Moreover, regardless of the initial precision \( h_1 \) about the worker’s ability, \( \{\mu_t\} \) converges monotonically to the unique steady–state \( \mu_\infty \) of (8); that is, \( \mu_t \) increases
monotonically to $\mu_\infty$ if $\mu_1 < \mu_\infty$ and $\mu_t$ decreases monotonically to $\mu_\infty$ if $\mu_1 > \mu_\infty$. Note that $\mu_\infty \in (0, 1)$ for all $\xi > 0$. We also known from Holmström that if

$$f_k(\mu_t) = (1 - \mu_t) \prod_{s=1}^{k} \mu_{t+s},$$

then $f_k$ is decreasing in $\mu_t$ for all $k \geq 0$. Notice that $f_0(\mu_t) = 1 - \mu_t$.

For each period $t$, the equilibrium condition (7) defines $a^*_t$ as a function of $\mu_t$, and thus as a function of the period–$t$ precision about the worker’s ability. We have the following result.

**Proposition 2.** For each $t \leq T - 1$, the worker’s choice of effort in period $t$ is strictly decreasing in the precision about his ability. Moreover, holding the precision about the worker’s ability constant, the worker’s choice of effort decreases strictly with his age.

**Proof:** Since the right–hand side of (8) is increasing in $\mu_t$, the first result follows from the fact that $f_k$ is decreasing in $\mu_t$ for all $k \geq 0$ and $\mu_t$ is strictly increasing in $h_t$. The second result follows immediately from (7).

Thus, as in Holmström, the smaller the precision about the worker’s ability, the greater his incentive to exert effort: the more there is uncertainty about the worker’s ability, the greater is the scope to for the worker to manipulate the market’s assessment of his ability. We refer to this effect on career concern incentives as the precision effect. Moreover, since the worker is finitely lived, his gain from influencing the market’s perception about his ability decreases with his age. We refer to this effect on career concern incentives as the tenure effect. Notice that $k'$ plays no role in the proof of Proposition 2, and so this result is also valid in a finite–horizon version of Holmström. The tenure effect is absent in Holmström’s environment since life cycle considerations are absent when the worker is infinitely lived.

The tenure effect always acts to decrease the worker’s choice of effort over time. The sign of the precision effect depends on whether the precision about the worker’s ability increases or decreases with his age. Consider the case where $\mu_1 \leq \mu_\infty$. Since $h/(h + h_\infty)$ is strictly increasing in $h$, $\mu_1 < \mu_\infty$ implies that the precision about the worker’s ability increases strictly to the steady–state precision $h_\infty = h_\xi \mu_\infty/(1 - \mu_\infty)$. Thus, when $\mu_1 \leq \mu_\infty$, either the precision effect works in the same direction as the tenure effect or the precision effect is absent (when the initial precision is the steady–state precision). In this case, the worker’s
choice of effort is strictly decreasing over time. In fact, the same is true even when \( \mu_1 > \mu_\infty \), as long as \( \mu_1 \) is close enough to \( \mu_\infty \) for the tenure effect to dominate the precision effect. To summarize, we have the following result as a consequence of Proposition 2.

**Proposition 3.** There exists \( 1 > \overline{\mu} > \mu_\infty \) such that if \( \mu_1 \leq \overline{\mu} \), then the equilibrium choice of effort is strictly decreasing over time.

In particular, the worker’s choice of effort is strictly decreasing over time when there is no accumulation of human capital \((k' \equiv 0)\) even when the uncertainty about the worker’s ability is fully replenished in each period \((\mu_1 = \mu_\infty)\). This stands in sharp contrast to the infinite horizon case considered by Holmström, where the worker’s choice of effort is constant over time in this case.\(^3\) As discussed above, the reason for this is that even though the precision effect is absent when \( \mu_1 = \mu_\infty \), the tenure effect is still present, and so the worker’s incentive to exert effort decreases with his age. This makes the case where \( \mu_1 = \mu_\infty \) a natural benchmark to study the optimal provision of explicit incentives in the presence of career concerns and is the reason why we introduced shocks to ability in our environment.

We now discuss the efficiency of the worker’s equilibrium behavior. For this, let \( a_t^{**} \) denote the efficient choice of effort in period \( t \). It is straightforward to see that \( a_t^{**} \) satisfies

\[
1 + \sum_{s=t+1}^{T} \delta^{s-t}k'(a) = g'(a). \tag{9}
\]

Notice that \( a_t^{**} \) is strictly decreasing over time when there is accumulation of human capital. Indeed, the human capital a worker accumulates early in his life has greater overall value than the human capital he accumulates later in life. We have the following result.

**Proposition 4.** There exists \( \underline{\mu} < \mu_\infty \) such that the accumulation of human capital is inefficiently low if \( \mu_1 > \underline{\mu} \).

**Proof:** Suppose that \( \mu_1 \geq \mu_\infty \). Since \( f_k(\mu_t) \) is decreasing in \( \mu_t \) for all \( k \geq 0 \), we have that

\[
\sum_{s=t+1}^{T} \delta^{s-t} \left[ (1 - \mu_t) \prod_{\tau=t+1}^{s-1} \mu_\tau + \sum_{q=t+1}^{s-1} (1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_\tau k'(a) \right] \\
\leq (1 - \mu_\infty) \sum_{s=t+1}^{T} \delta^{s-t} \left[ \mu^{s-t-1}_\infty + \sum_{q=t+1}^{s-1} \mu^{s-q-1}_\infty k'(a) \right] \leq 1 - \mu^{T-t}_\infty + \sum_{s=t+1}^{T} \delta^{s-t}[1 - \mu^{s-t-1}_\infty]k'(a).
\]

\(^3\)Our environment also differs from Holmström (1999) in our assumption that workers are risk–averse. It is clear from the analysis above that risk–aversion plays no role in our results about career concern incentives.
From (7), (9), and the above inequality, we can then conclude that $a_{t}^{**} > a_{t}^{*}$ for all $t \leq T$. By continuity, $a_{t}^{**} > a_{t}^{*}$ for all $t \leq T$ even if $\mu_{1} < \mu_{\infty}$, as long as $\mu_{1}$ is close enough to $\mu_{\infty}$.

The intuition for Proposition 4 is as follows. At any point in time, except in the last period, the worker’s career concern incentive to exert effort is increasing in the uncertainty about his ability. If the initial uncertainty about the worker’s ability is not high enough ($\mu_{1}$ is not small enough), the worker’s choice of effort will always be inefficiently low. However, as the following example shows, it is possible that the worker’s choice of effort is inefficiently high early on if the initial uncertainty about his ability is high enough. We return to this point in the next section. Assume that $\delta = 1$, $\xi \approx 0$, and $\mu_{1} = 0$ (i.e., $h_{1} = 0$). Equation (8) then implies that $\mu_{s} \approx (s - 1)/s$, in which case (7) reduces (approximately) to

$$ \sum_{t=s+1}^{T} \left[ \frac{1}{t-1} + \frac{t-s-1}{t-1} k'(a) \right] = g'(a). $$

Assume now that $k(a) = \gamma a$ and set $s = 1$. Since $\sum_{t=2}^{T} 1/(t-1) > 1$ as long as $T \geq 3$, the above condition implies $a_{1}^{*} > a_{1}^{**}$ if $T \geq 3$ when $\gamma$ is sufficiently small.

3 Linear Contracts

We now consider the case where in every period the firms in the market offer short–term linear contracts to the worker, so that the worker’s wage in period $t$ is $w_{t} = c_{t} + b_{t}y_{t}$, where $b_{t}$ is the piece rate and $c_{t}$ is base wage in the contract offered to the worker.\footnote{Gibbons and Murphy (1992) show that the environment where long–term contracts are feasible, but equilibria must be renegotiation–proof, is outcome equivalent to the environment where only short–term contracts are possible.} We first show that there exists a unique equilibrium and provide a complete characterization of this equilibrium. We then study how the presence of human capital accumulation affects the time pattern of incentives (total and explicit).

3.1 Equilibrium

A behavior strategy $\sigma$ for the worker is the same as in the benchmark model. A strategy for the market is now a sequence $\omega = \{\omega_{t}\}_{t=1}^{T}$, such that $\omega_{t}(y_{t})$ is the linear contract offered
to the worker in period \( t \) when his output history is \( y^t \); we denote the base wage and piece rate in \( \omega_t(y^t) \) by \( \gamma_t(\omega, y^t) \) and \( \beta_t(\omega, y^t) \), respectively. We say the worker’s period–\( t \) piece rate under \( \omega \) is uncontingent if \( \beta_t(\omega, y^t) \) is the same for all \( y^t \). The worker is still paid his expected output in every period. Thus, if the market expects the worker to follow the strategy \( \sigma \) and the worker’s period–\( t \) output history is \( y^t \), then \( w_t = c_t + b_t y_t \) with \( c_t = (1 - b_t)\mathbb{E}[y_t|\sigma, y^t] \).

Let \( \Pi(\sigma, \omega) \) be the lifetime payoff to the worker when he follows the strategy \( \sigma \) and the market’s strategy is \( \omega \). Now let \( \Theta \) be the set of all pairs \( (\sigma, \omega) \) such that \( \sigma \) is sequentially rational given \( \omega \) and \( \beta_t(\omega, y^t) \) and \( \gamma_t(\omega, y^t) \) are such that \( \gamma_t(\omega, y^t) = (1 - \beta_t(\omega, y^t))\mathbb{E}[y_t|\sigma, y^t] \).

Notice that \( \Theta \) depends on the (common) prior about the worker’s ability and on the worker’s initial human capital. For each output \( y \) and strategy \( \sigma \), let \( \Psi(\sigma, y) \) be the worker’s reputation in \( t = 2 \) when he plays according to \( \sigma \) and produces \( y \) in \( t = 1 \). Moreover, for each output \( y \) and pair \( (\sigma, \omega) \), let \( \sigma_y = \{\sigma_{t,y}\}_{t=1}^{T-1} \) and \( \omega_y = \{\omega_{t,y}\}_{t=1}^{T-1} \) be such that \( \sigma_{t,y}(y^t) = \sigma_{t+1}(y, y^t) \) and \( \omega_{t,y}(y^t) = \omega_{t+1}(y, y^t) \). We define equilibria inductively in the worker’s lifetime \( T \). For this, it is convenient to assume that \( T = 1 \) is possible.

**Definition 2.** Suppose that \( T = 1 \). The pair \( (\sigma^*, \omega^*) \) is an equilibrium if it belongs to \( \Theta \) and \( \Pi(\sigma^*, \omega^*) \geq \Pi(\sigma, \omega) \) for all \( (\sigma, \omega) \) in \( \Theta \). Suppose now that \( T \geq 2 \). The pair \( (\sigma^*, \omega^*) \) is an equilibrium if:

1. The pair \( (\sigma^*, \omega^*) \) is an element of \( \Theta \);
2. For each output \( y \), \( (\sigma^*_y, \omega^*_y) \) is an equilibrium when the worker’s lifetime is \( T - 1 \), the prior about the worker’s ability is \( \Psi(\sigma^*, y) \), and the worker’s initial human capital is \( k_1 + k(\sigma^*_1(y^1)) \);
3. \( \Pi(\sigma^*, \omega^*) \geq \Pi(\sigma, \omega) \) for all \( (\sigma, \omega) \) in \( \Theta \) that satisfy condition 2.

As in Section 2, we restrict attention to equilibria in which the worker follows a pure strategy. In what follows, it is more convenient to work with variances instead of precisions. For this, let \( \sigma^2_\varepsilon = \frac{1}{h_\varepsilon} \) be the variance of the noise terms \( \varepsilon_t \), \( \sigma^2_\eta = \frac{1}{h_\eta} \) be the variance of the shock terms \( \eta_t \), and \( \sigma^2_t = \frac{1}{h_t} \) be the variance of the worker’s belief about his ability in period \( t \). Moreover, let \( \Sigma^2_t = \sigma^2_t + \sigma^2_\varepsilon \). We have the following result, the proof of which is in the Appendix.
Proposition 5. Suppose that either $k$ is linear or $k'(a) \leq 1 + r\sigma^2 g''(a)$ for all $a$. There exists a unique equilibrium, which is uncontingent. Let $a_t^*$ be the worker’s choice of effort in period $t$ and $b_t^*$ be the piece rate in period $t$. The pair $(a_t^*, b_t^*)$ is determined as follows. Let

$$B_t^* = b_t^* + \sum_{s=t+1}^{T} \delta^{s-t}(1 - b_s^*)(1 - \mu_t) \prod_{\tau=t+1}^{s-1} \mu_\tau$$

and denote by $a_t^*(B)$ the unique solution to

$$g'(a) = B + \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^* k'(a).$$

Then $a_t^* = a_t^*(B_t^*)$, where $B_t^*$ is the unique solution to

$$H_t(B) = 1 - B - r \left[ g''(a_t^*(B)) - k''(a_t^*(B)) \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^* + \right. \left[ \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^* \delta^{\tau-t} B_\tau^* \right] \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^* - \left.\right] k'(a_t^*(B)) \sum_{\tau=t+1}^{T} \delta^{\tau-t}(1 - B_\tau^*) = 0.$$

3.2 The Time Pattern of Incentives

We first analyze total incentives. Then we discuss the case without accumulation of human capital. After that we discuss the lifetime profile of total incentives in the linear case.

3.3 Total Incentives (General Case)

Suppose that $g''(0) > 0$. From Proposition 5, we have that

$$B_t^* = \frac{1 + k'(a_t^*) \sum_{\tau=t+1}^{T} \delta^{\tau-t}}{1 + r \sum_{t+1}^{T} \left[ g''(a_t^*) - k''(a_t^*) \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^* \right]}$$

Let $S_t^* = \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^*$. Then

$$B_t^* + \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^* = \underbrace{\frac{1 + k'(a_t^*) \sum_{\tau=t+1}^{T} \delta^{\tau-t}}{1 + r \sum_{t+1}^{T} \left[ g''(a_t^*) - k''(a_t^*) S_{t+1}^* \right]} \underbrace{b_t^*}_{b_t^*(a_t^*)} \left[ \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^* \right]}_{1 - \xi_t(a_t^*)}.$$
which implies that
\[ S_t^* = \delta \left[ b_t^*(a_t^*) + (1 - \xi_t^*(a_t^*))S_{t+1}^* \right] \tag{10} \]

Also notice that
\[ \hat{B}_t^* = B_t^* + k'(a_t^*) \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_{\tau}^* = b_t^*(a_t^*) + [k'(a_t^*) - \xi_t^*(a_t^*)] S_{t+1}^*. \tag{11} \]

We now show that total incentives are always positive if \( \sigma_T^2/\Sigma_T^2 \leq k'(0) \leq 1 + r\sigma_T^2g''(0) \) for all \( t \). First notice that \( S_T^* = \delta B_T^* \) and
\[ B_T^* = \frac{1}{1 + r\Sigma_T^2g''(a_T^*)} > 0. \]

Thus, \( a_T^* > 0 \) and \( S_T^* > 0 \). Suppose then, by induction, that \( a_T^* \) and \( S_T^* > 0 \) for all \( \tau \geq t + 1 \), where \( t \in \{1, \ldots, T-1\} \). We want to show that \( S_t^* > 0 \) and \( a_t^* > 0 \). First notice that
\[ k'(a) - \xi_t^*(a) = \frac{r(\Sigma_T^2k'(a) - \sigma_T^2) [g''(a) - k''(a)S_{t+1}^*]}{1 + r\Sigma_T^2 [g''(a) - k''(a)S_{t+1}^*]}. \]

Since \( S_{t+1}^* > 0 \) (and so the denominator of \( b_t^*(a) \) is positive, which implies that \( b_t^*(a) > 0 \)) and \( k'(a) \geq \sigma_T^2/\Sigma_T^2 \) for all \( a \leq 0 \), we then have that
\[ b_t^*(a) + [k'(a) - \xi_t^*(a)] S_{t+1}^* > 0, \]

for all \( a \leq 0 \), so that \( a_t^* > 0 \) by (11); recall that \( g'(0) = 0 \) and \( g'(a) < 0 \) for \( a < 0 \) (the cost of stealing output increases with the amount of output stolen). Now notice that \( \xi_t^*(a_t^*) \leq 1 \) if, and only if,
\[ r\sigma_T^2 [g''(a_t^*) - k''(a_t^*)S_{t+1}^*] + k'(a_t^*) < 1 + r\Sigma_T^2 [g''(a_t^*) - k''(a_t^*)S_{t+1}^*]; \]

that is, if and only if,
\[ k'(a_t^*) < 1 + r\sigma_T^2 [g''(a_t^*) - k''(a_t^*)S_{t+1}^*]. \]

Since \( S_{t+1}^* > 0 \), \( a_t^* > 0 \), and \( g'' \geq 0 \), we have that \( g''(a_t^*) - k''(a_t^*)S_{t+1}^* \geq g''(0) \) and \( k'(a_t^*) < k'(0) \). Thus, the last inequality is satisfied since \( k'(0) \leq 1 + r\sigma_T^2g''(0) \). Thus, \( S_t^* > 0 \) by (10).
3.4 Total Incentives in the Linear Case

If \( k(a) = \gamma a \), then total incentives are always positive as long as \( \sigma^2_t / \Sigma^2_t \leq \gamma \leq 2[1 + r\sigma^2_\varepsilon g''(0)] \) for all \( t \). First notice that (11) reduces to

\[
\hat{B}^*_t = b^*_t + [\gamma - \xi^*_t] S^*_t + 1,
\]

where we omit the dependence of \( b^*_t \) and \( \xi^*_t \) on \( a^*_t \) for ease of notation. As before, \( a^*_T > 0 \) and \( S^*_T = \delta B^*_T > 0 \). Suppose then, by induction, that \( S^*_\tau > 0 \) and \( a^*_\tau > 0 \) for all \( \tau \geq t + 1 \), with \( t \in \{1, \ldots, T - 1\} \). We want to show that \( S^*_t > 0 \) and \( a^*_t > 0 \). For this notice that

\[
\xi^*_t = \frac{r\sigma^2_\varepsilon g''(a^*_t)}{1 + r\Sigma^2_\varepsilon g''(a^*_t)} < \gamma \quad \text{and} \quad b^*_s = \frac{1 + \gamma \sum_{\tau=t+1}^T \delta^{\tau-s}}{1 + r\Sigma^2_\varepsilon g''(a^*_s)} > 0
\]

for all \( s \). Since \( S^*_{t+1} > 0 \), (12) implies that \( \hat{B}^*_t > 0 \), and so \( a^*_t > 0 \). If \( \xi^*_t \leq 1 \), equation (10) implies that \( S^*_t > 0 \). Suppose then that \( \xi^*_t > 1 \)—this happens when \( \gamma \geq 1 + r g''(a^*_s) \sigma^2_\varepsilon \), a situation compatible with \( \gamma \leq 2[1 + r g''(0) \sigma^2_\varepsilon] \). Since \( S^*_{t+2} \geq 0 \) by assumption, equation (10) implies that \( S^*_{t+1} \leq \delta b^*_{t+1} \), and so (since \( 1 - \xi^*_t \leq 0 \)), \( S^*_t \geq \delta [b^*_t + (1 - \xi^*_t) \delta b^*_{t+1}] \). Thus, \( S^*_t \geq 0 \) if \( b^*_t + (1 - \xi^*_t) \delta b^*_{t+1} \geq 0 \). Now observe that

\[
\frac{1 - \xi^*_t}{1 + r g''(a^*_s) \Sigma^2_\varepsilon} = \frac{1}{1 + r g''(a^*_s) \Sigma^2_\varepsilon} \left( \frac{1 + r g''(a^*_s) \sigma^2_\varepsilon - \gamma}{1 + r g''(a^*_s) \Sigma^2_\varepsilon} \right) \geq \frac{-1}{1 + r g''(a^*_s) \Sigma^2_\varepsilon}
\]

if, and only if,

\[
\gamma \leq 2 + r g''(a^*_s) \sigma^2_\varepsilon + r g''(a^*_t+1) \Sigma^2_\varepsilon,
\]

which is satisfied since \( a^*_s, a^*_t+1 \geq 0, g'' \geq 0, \) and \( \gamma \leq 2[1 + r \sigma^2_\varepsilon g''(0)] \). Hence,

\[
b^*_t + (1 - \xi^*_t) \delta b^*_{t+1} \geq \frac{1}{1 + r g''(a^*_s) \Sigma^2_\varepsilon} \left[ 1 + \gamma \sum_{\tau=t+1}^T \delta^{\tau-t} - \delta - \gamma \sum_{\tau=t+2}^T \delta^{\tau-t} \right] \geq 0,
\]

the desired result.

Notice that the restriction that \( k'(0) \leq 1 + r \sigma^2_\varepsilon g''(0) \) (or that \( \gamma \leq 2[1 + r \sigma^2_\varepsilon g''(0)] \) in the linear case) is easier to satisfy when \( r \) is large. This is intuitively plausible. Since the worker is risk–averse, he dislikes variation in his equilibrium choice of effort. Since this choice is always positive in the last period, period \( T \), a very risk–averse worker would rather exert positive effort in all periods (and avoid too much variation in his choices of effort over time).
This begs the question of whether total incentives can ever be negative when \( k'(0) \) (or \( \gamma \) in the linear case) is large. The example that follows shows that this can be the case.

In addition to \( k(a) = \gamma a \), assume that \( g(a) = a^2 / 2 \), and consider the steady–state case. For ease of notation, let \( \Sigma^2 = \sigma_1^2 + \sigma_\varepsilon^2 \), so that \( \Sigma^2 \equiv \Sigma^2 \) and

\[
\xi_t^* \equiv \frac{r \sigma_1^2 + \gamma}{1 + r \Sigma^2}.
\]

It is clear from the argument that follows that what matters is for our conclusions is that \( r g''(a^*_t) \) is constant in equilibrium when \( g \) is quadratic. So, the specific (quadratic) form of \( g \) we use does not matter.

In this case, the equation determining \( B^*_t \) reduces to

\[
H_t(B^*_t) = 1 - B^*_t - r \left[ B^*_t \Sigma^2 + \sigma_1^2 \sum_{\tau=t+1}^{T} \delta^{\tau-t} B^*_\tau \right] + \gamma \sum_{\tau=t+1}^{T} \delta^{\tau-t} (1 - B^*_\tau) = 0.
\]

We claim that

\[
B^*_{T-q} = B^*_{T+1-q} + \frac{r \delta^q \kappa}{(1 + r \Sigma^2)^2} \left\{ 1 - \left[ \frac{r \sigma_1^2 + \gamma}{1 + r \Sigma^2} \right]^{q-1} \right\}.
\]  

(13)

where \( \kappa = \Sigma^2 \gamma - \sigma_1^2 \), which implies that

\[
B^*_{T-q} = B^*_{T} + \frac{r \kappa}{(1 + r \Sigma^2)^2} \sum_{s=1}^{q} \delta^s \left( 1 - \left[ \frac{r \sigma_1^2 + \gamma}{1 + r \Sigma^2} \right]^{s-1} \right) = B^*_{T} + \frac{r \kappa}{(1 + r \Sigma^2)^2} \sum_{s=1}^{q} \delta^s (1 - \xi)^{s-1}.
\]  

(14)

First notice that the guess is correct when \( q = 1 \). Suppose then, by induction, that the guess is correct for all \( q' \leq q \), where \( q \in \{1, \ldots, T-2\} \), and let

\[
B = B^*_{T} + \frac{r \kappa}{(1 + r \Sigma^2)^2} z(q + 1),
\]

where \( z(q) = \sum_{s=1}^{q} \delta^s (1 - \xi)^{s-1} \). We want to show that \( H_{T-q-1}(B) = 0 \). For this, notice that

\[
H_{T-q-1}(B) = 1 - B - r \left[ B \Sigma^2 + \sigma_1^2 \sum_{\tau=T-q}^{T} \delta^{\tau-T+q+1} B^*_\tau \right] + \gamma \sum_{\tau=T-q}^{T} \delta^{\tau-T+q+1} (1 - B^*_\tau)
\]

\[
= 1 - \left( 1 + r \Sigma^2 \right) B^*_{T} - (1 + r \Sigma^2) \frac{r \kappa}{(1 + r \Sigma^2)^2} z(q + 1)
\]

\[-r \sigma_1^2 \sum_{\tau=T-q}^{T} \delta^{\tau-T+q+1} B^*_\tau + \gamma \sum_{\tau=T-q}^{T} \delta^{\tau-T+q+1} (1 - B^*_\tau).
\]
Now observe, by the induction hypothesis (and (14)), that
\[ \sum_{\tau=T-q}^{T} \delta^{q+1}(T-\tau) B^*_\tau = \frac{1}{1 + r\sum^2} \sum_{\tau=T-q}^{T} \delta^{q+1}(T-\tau) + \frac{r\kappa}{(1 + r\sum^2)^2} \sum_{\tau=T-q}^{T} \delta^{q+1}(T-\tau) \]
and that
\[ \sum_{\tau=T-q}^{T} \delta^{q+1}(1 - B^*_\tau) = \frac{r\sum^2}{1 + r\sum^2} \sum_{\tau=T-q}^{T} \delta^{q+1}(T-\tau) - \frac{r\kappa}{(1 + r\sum^2)^2} \sum_{\tau=T-q}^{T} \delta^{q+1}(T-\tau). \]
Hence,
\[ H_{T-q-1}(B) = \frac{r\kappa}{1 + r\sum^2} \left\{ \sum_{\tau=T-q}^{T} \delta^{q+1}(1 - \delta(1 - \xi)T-\tau) \right\} = \frac{r\kappa}{1 + r\sum^2} \left\{ \sum_{\tau=T-q}^{T} \delta^{q+1}(1 - \delta(1 - \xi)T-\tau) \right\}. \]

We claim that the term in brackets in the last equation is zero, so that \( B = B^*_T \). Indeed,
\[ \sum_{s=1}^{q+1} \delta^s(1 - \xi)^{s-1} = \frac{\delta[1 - \delta q^2 + \gamma]}{1 - \delta(1 - \xi)} \]
and
\[ 1 - \xi \sum_{s=1}^{T-\tau} \delta^s(1 - \xi)^{s-1} = 1 - \xi \frac{[1 - \delta^q(1 - \xi)^{T-\tau}]}{1 - \delta(1 - \xi)} = 1 - \delta + \delta \xi \delta^q(1 - \xi)^{T-\tau}, \]
so that
\[ \sum_{\tau=T-q}^{T} \delta^{q+1}(1 - \xi)^{s-1} = \frac{1 - \delta}{1 - \delta(1 - \xi)} \sum_{\tau=T-q}^{T} \delta^{q+1}(1 - \xi)^{T-\tau} + \frac{\delta \xi \sum_{\tau=T-q}^{T} \delta^{q+1}(1 - \xi)^{T-\tau}}{1 - \delta(1 - \xi)} = \frac{\delta[1 - \delta q^2 + \gamma]}{1 - \delta(1 - \xi)} = \sum_{s=1}^{q+1} \delta^s(1 - \xi)^{s-1}. \]

Let us now investigate what happens to total incentives when \( \gamma \) is large. First notice, by
that
\[ B_{T-q}^* = B_{T-q}^* + \gamma(\delta B_{T-q+1}^* + \cdots + \delta q B_T^*) \]
\[ = (1 + \gamma\delta + \cdots + \gamma\delta^q)B_T^* + \frac{r\kappa\gamma}{(1 + r\Sigma^2)^2} \sum_{s=T-q+1}^T \delta^{T-s}q \sum_{s=1}^q \delta^s(1 - \xi)^s-1 \]
\[ + \frac{r\kappa}{(1 + r\Sigma^2)^2} \sum_{s=1}^q \delta^s(1 - \xi)^s-1. \]

In particular, when \( \delta = 1, \)
\[ \hat{\delta}_{T-q}^* = \frac{1 + q\gamma}{1 + r\Sigma^2} + \frac{r\kappa\gamma}{(1 + r\Sigma^2)^2} \left[ \frac{q\xi - 1 + (1 - \xi)^q}{\xi^2} \right] + \frac{r\kappa}{(1 + r\Sigma^2)^2} \left[ \frac{1 - (1 - \xi)^q}{\xi} \right]. \]

Now observe that
\[ \lim_{\gamma \to \infty} \frac{\gamma}{\xi} = 1 + r\Sigma^2 \quad \text{and} \quad \lim_{\gamma \to \infty} \frac{\kappa}{\gamma} = \Sigma^2. \]

Hence, when \( \gamma \) is large,
\[ \frac{r\kappa\gamma}{(1 + r\Sigma^2)^2} \left[ \frac{q\xi - 1 + (1 - \xi)^q}{\xi^2} \right] + \frac{r\kappa}{(1 + r\Sigma^2)^2} \left[ \frac{1 - (1 - \xi)^q}{\xi} \right] \approx \frac{r\Sigma^2\gamma}{(1 + r\Sigma^2)^2} \left[ \frac{(1 + r\Sigma^2)[q\xi - 1] + 1 + r\Sigma^2(1 - \xi)^q}{\xi} \right], \]
and so
\[ \hat{\delta}_{T-q}^* \approx \frac{1}{1 + r\Sigma^2} + \frac{\gamma}{1 + r\Sigma^2} \left\{ q + \frac{r\Sigma^2}{1 + r\Sigma^2} \left[ \frac{(1 + r\Sigma^2)[q\xi - 1] + 1 + r\Sigma^2(1 - \xi)^q}{\xi} \right] \right\}. \]

To finish, observe that
\[ \lim_{\xi \to \infty} \frac{(1 + r\Sigma^2)[q\xi - 1] + 1 + r\Sigma^2(1 - \xi)^q}{\xi} = \lim_{\xi \to \infty} q(1 + r\Sigma^2) - r\Sigma^2q(1 - \xi)^{q-1} = -\infty \]
when \( q \geq 2 \) and \( q - 1 \) is even. Thus total incentives become negative when \( \gamma \) is large enough if \( q \geq 2 \) and \( q \) is odd.

Since \( \xi_t^* \equiv \xi < \gamma, \) equation (12) implies that total incentives are negative in period \( t \) only if \( S_{t+1}^* < 0. \) But, then, equation (10) implies that \( S_t^* > 0, \) given that total incentives can only be negative when \( \xi > 1. \) Thus, by (12), if total incentives are negative in some period \( t, \) they are positive in period \( t - 1, \) that is, total incentives oscillate between being positive and negative. This is in accord with the intuition, given at the very beginning of the example, for why total incentives are always positive if the worker is sufficiently risk averse. Finally, it
is easy to see that, regardless of $\gamma$, total incentives eventually become positive in every period as $g''(a)$ approaches zero. This result is not surprising since reducing $g''(a)$ is equivalent to making the worker less risk averse.

3.5 No Human Capital Benchmark

Let us consider the no human capital benchmark. In this case, the equation for $B_t^*$ reduces to

$$H_t(B_t^*) = 1 - B_t^* - r g''(a_t^*) \left[ \Sigma_t^2 B_t^* + \sigma_t^2 \sum_{\tau=t+1}^{T} \delta^{\tau-t} B_\tau^* \right] = 0,$$

(15)

where $a_t^*$ is the unique solution to $g'(a_t^*) = B_t^*$.

Notice that (15) is formally equivalent to equation (A5) in Gibbons and Murphy. The only difference is that now the terms $\sigma_t^2$ evolve differently. Suppose then that $\sigma_1^2 \geq \sigma_\infty^2$, so that $\sigma_t^2 \geq \sigma_{t+1}^2$ for all $t \geq 1$, i.e., the uncertainty about the worker’s ability is (weakly) decreasing over time. We can then use the argument in Gibbons and Murphy (Lemmas 1 to 3 in the Appendix) to prove that $B_t^*$ is strictly increasing in $t$. We cannot use Lemma 4 to conclude that explicit incentives are strictly increasing, though, as the expression for $B_t$ as a function of the piece rates is now different. We need to establish this result in our case.

First notice that $B_T^* = b_T^*$ and that

$$B_t^* = b_t^* + \sum_{\tau=t+1}^{T} \delta^{\tau-t}(1 - b_t^*)(1 - \mu_t)\mu_{t+1} \cdots \mu_{\tau-1},$$

where we adopt the convention that $\mu_{t+1} \cdots \mu_{\tau-1} = 1$ when $\tau = t + 1$. Since $b_{T-1}^* < B_{T-1}^*$ and $B_{T-1}^* < B_T^*$, we then have that $b_{T-1}^* < b_T^*$. Suppose now, by induction, that $b_t^* < b_{t+1}^*$

---

5Recall that we denote the variance of the prior belief about the worker’s ability by $\sigma_1^2$, while Gibbons and Murphy denote this variance by $\sigma_0^2$. With our convention, $\Sigma_t^2 = \sigma_t^2 + \sigma_\infty^2$.

6In Gibbons and Murphy the variance $\sigma_t^2$ is strictly decreasing. It is immediate to see that the same result holds if the variances $\sigma_t^2$ are constant over time.
for all $\tau \geq t + 1$, with $t \in \{2, \ldots, T - 1\}$. Moreover, let $f_k(\mu_s) = (1 - \mu_s)\mu_{s+1} \cdots \mu_{s+t}$. Then,

$$b_{t+1}^* - b_t^* = B_{t+1}^* - B_t^* + \sum_{\tau=t+1}^{T} \delta^{\tau-t}(1 - b_t^*)f_{\tau-t-1}(\mu_t) - \sum_{\tau=t+2}^{T} \delta^{\tau-t-1}(1 - b_t au^*)f_{\tau-t-2}(\mu_{t+1})$$

$$> \sum_{\tau=t+1}^{T-1} \delta^{\tau-t}(1 - b_t^*)f_{\tau-t-1}(\mu_t) + \delta^{T-t}(1 - b_T^*)f_{T-t-1}(\mu_t)$$

$$- \sum_{\tau=t+1}^{T-1} \delta^{\tau-t}(1 - b_{t+1}^*)f_{\tau-t-1}(\mu_{t+1})$$

$$> \sum_{\tau=t+1}^{T-1} \delta^{\tau-t}(1 - b_t^*)f_{\tau-t-1}(\mu_t) - \sum_{\tau=t+1}^{T-1} \delta^{\tau-t}(1 - b_{t+1}^*)f_{\tau-t-1}(\mu_{t+1}),$$

since $b_T^* < 1$ and $B_t^*$ is strictly increasing. Now observe, by the induction hypothesis, that $1 - b_T^* > 1 - b_{t+1}^* > 0$ for all $\tau \geq t + 1$. We also know, from Hölmstrom, that $f_k(\mu_t)$ is decreasing in $\mu_t$. Since $\sigma_1^2 \geq \sigma_\infty^2$ implies that $\mu_1 \leq \mu_\infty$ and $\mu_t$ increases monotonically to $\mu_\infty$, we also have that $f_{\tau-t-1}(\mu_t) \geq f_{\tau-t-1}(\mu_{t+1})$ for all $\tau \geq t + 1$. Hence, $b_{t+1}^* > b_t^*$. We can then conclude, as in Gibbons and Murphy, that $b_t^*$ is strictly increasing in $t$.

Notice, in particular, that these results are true in the steady–state. The reason is that the equilibrium contract trades off efficiency, which dictates a constant choice of effort over time, with insurance, due to the noise in performance and the presence of uncertainty about ability. Even when uncertainty about ability is constant over time, the insurance motive is greater when the worker is younger. Indeed, given the same uncertainty about the worker’s ability, a longer lifetime implies a greater variation in lifetime payoffs (holding everything else constant). The driving force behind the results in Gibbons and Murphy is not the uncertainty effect, but the tenure effect.

### 3.6 Lifetime Profile of Incentives (Linear Case)

Suppose we are on the steady–state (i.e., $\sigma_1^2 = \sigma_\infty^2$) and recall that

$$\hat{B}_t^* = B_t^* + \sum_{\tau=t+1}^{T} \delta^{\tau-t}B_{\tau}^*\gamma$$

20
denotes total incentives in period $t$ ($a_i^*$ is the solution to $g'(a_i^*) = \hat{B_i}$). From the equation $H_t(B_t) = 0$ we obtain that

$$
\hat{H}_t(\hat{B}_t^*) := 1 - \hat{B}_t^* - r g''(a_t^*) \left[ \hat{B}_t^* \Sigma_1^2 - (\Sigma_1^2 \gamma - \sigma_1^2) \sum_{\tau=t+1}^T \delta^{\tau-t} B_t^* \gamma \right] + \gamma \sum_{\tau=t+1}^T \delta^{\tau-t} = 0; \quad (16)
$$

notice that in the above equation we treat $B_t^*$, with $\tau \geq t + 1$, as constants. Now observe (just to be sure) that

$$
\hat{B}_t^* > \hat{B} \iff \hat{B} - \sum_{\tau=t+1}^T \delta^{\tau-t} B_t^* \gamma < B_t^* \iff H_t \left( \hat{B} - \sum_{\tau=t+1}^T \delta^{\tau-t} B_t^* \gamma \right) = \hat{H}_t(\hat{B}) > 0.
$$

We claim that if $\sigma_1^2/\Sigma_1^2 = \sigma_\infty^2/\Sigma_\infty^2 \leq \gamma \leq 1$, then total incentives are strictly decreasing. Notice first that

$$
\hat{H}_{T-1}(\hat{B}_{T-1}^*) = 1 - \hat{B}_{T-1}^* - r g''(a_{T-1}^*) \Sigma_1^2 \hat{B}_{T-1}^* + \gamma \delta = \gamma \delta > 0,
$$

since $\hat{B}_T = B_T^*$. Hence, $\hat{B}_{T-1}^* > \hat{B}_{T-1}^*$. Since $\gamma \leq 1$ (and $B_T^* \geq 0$), we then have that

$$
B_{T-1}^* + \delta B_{T}^* \geq B_{T-1}^* + \delta B_T^* \gamma = \hat{B}_{T-1}^* > \hat{B}_{T} = B_T^*.
$$

Suppose then, by induction, that

$$
\sum_{\tau=t+1}^T \delta^{\tau-t} B_t^* \geq \sum_{\tau=t+2}^T \delta^{\tau-t-1} B_t^* \geq 0
$$

for some $t \in \{2, \ldots, T-2\}$ (this is true when $t = T-2$). Now observe that

$$
\begin{align*}
\hat{H}_t(\hat{B}_{t+1}^*) &= 1 - \hat{B}_{t+1}^* - r g''(a_{t+1}^*) \left[ \hat{B}_{t+1}^* \Sigma_1^2 - (\Sigma_1^2 \gamma - \sigma_1^2) \sum_{\tau=t+1}^T \delta^{\tau-t} B_t^* \gamma \right] + \gamma \sum_{\tau=t+1}^T \delta^{\tau-t} \\
&= -r g''(a_{t+1}^*) \left\{ (\sigma_1^2 - \Sigma_1^2 \gamma) \left[ \sum_{\tau=t+1}^T \delta^{\tau-t} B_t^* - \sum_{\tau=t+2}^T \delta^{\tau-t-1} B_t^* \right] \right\} + \gamma \delta^{T-t}.
\end{align*}
$$

Since $\gamma > \sigma_1^2/\Sigma_1^2$ by assumption, we then have that $\hat{H}_t(\hat{B}_{t+1}^*) > 0$. Thus, $\hat{B}_t^* > \hat{B}_{t+1}^*$. To finish, observe that

$$
\delta B_t^* + \delta^2 B_{t+1}^* + \cdots + \delta^{T-t+1} B_T^* = \delta [B_t^* + \delta B_{t+1}^* + \cdots + \delta^{T-t} B_T^*] \\
= \delta [\hat{B}_t^* + (\delta B_{t+1}^* + \cdots + \delta^{T-t} B_T^*) (1 - \gamma)] \\
> \delta [\hat{B}_t^* + (\delta B_{t+2}^* + \cdots + \delta^{T-t-1} B_T^*) (1 - \gamma)] \\
= \delta [B_{t+1}^* + \delta B_{t+2}^* + \cdots + \delta^{T-t-1} B_T^*].
$$
We can then conclude, by induction, that total incentives are indeed strictly decreasing.

When $\sigma^2_1 \downarrow 0$, the expression for the $B^*_t$ implies that $B^*_t$ converges to $b_t^*$ by definition of $B^*_t$. Thus, in the steady–state, explicit incentives are strictly decreasing in the linear case if $\sigma^2_1 / \Sigma^2_1 \leq \gamma \leq 1$ as long as $\sigma^2_1$, and so the steady–state variance $\sigma^2_\infty$, is small enough. The usual continuity argument allows us to move a bit from the steady–state—the interesting direction, of course, is when $\sigma^2_1$ becomes greater than $\sigma^2_\infty$. These results about explicit incentives are not so interesting, since having $\sigma^2_1$ small requires us to have $\sigma^2_\infty$ small as well, unless we consider the case where variance about ability increases over time.

We conclude our discussion of the linear human capital accumulation case with a few remarks. Notice that if $\gamma = 1$, then $\Sigma^2_t \gamma - \sigma^2_t = \sigma^2_t$ for all $t \geq 1$. In this special case, the above argument also works outside of the steady–state regardless of whether the initial value of $\sigma^2_t$ is above or below $\sigma^2_\infty$. Now, if $\gamma < 1$, then $\Sigma^2_t \gamma - \sigma^2_t = \sigma^2_t (\gamma - 1) + \sigma^2_t$, which is decreasing in $\sigma^2_t$. Thus, the above argument remains valid even when $\sigma^2_t > \sigma^2_\infty$. So, total incentives are strictly decreasing as long as $\sigma^2_t \leq \sigma^2_\infty$, under the assumption that $\sigma^2_\infty / \Sigma^2_\infty \leq \gamma \leq 1$.

In addition, by continuity total incentives are strictly decreasing when $\sigma^2_t > \sigma^2_\infty$ as long as $\sigma^2_t$ is close enough to $\sigma^2_\infty$ (and $\sigma^2_\infty / \Sigma^2_\infty \leq \gamma \leq 1$). Certainly, wow close $\sigma^2_t$ needs to be to $\sigma^2_\infty$ depends on $T$. However, when $\sigma^2_t > \sigma^2_\infty$, for fixed $T$, total incentives are eventually decreasing as long as $\sigma^2_t$ is close enough to $\sigma^2_\infty$. Finally, we know from Subsection 1.2 that total incentives can be negative in the linear case when $\gamma$ is very large. In particular, since total incentives are always positive in the last period, total incentives need not be decreasing over time when $\gamma > 1$ even in the steady–state.

3.7 Lifetime Profile of Incentives (General Case)

We now consider total incentives in the general case ($k$ linear or non–linear). Suppose that $r = 0$. Then, $B^*_t = 1$, which implies that $S^*_t = \delta$. Suppose now that $S^*_{t+1} = \delta + \cdots + \delta^{T-t}$ for some $t \in \{1, \ldots, T-1\}$. Since $S^*_{t+1} > 0$, we have that $g''(a) - k''(a)S^*_{t+1} \geq 0$ for all $a$. Hence, $B^*_t \leq 1$, which implies that $a^*_t$ is bounded, and so $rg''(a^*_t) = 0$. This implies that $B^*_t = 1$, and so $S^*_t = \delta (1 + S^*_{t+1}) = \delta + \cdots + \delta^{T-t+1}$. Thus, $S^*_t = \delta + \cdots + \delta^{T-t+1}$ for all $t \geq 1$.  

22
Consequently, 

\[
\tilde{B}_t^* = B_t^* + \gamma S_{t+1}^* = 1 + \gamma (\delta + \cdots + \delta^{T-t}),
\]

which is decreasing in \(t\). In other words, total incentives are strictly decreasing in the limit case when \(r = 0\). By continuity, total incentives are strictly decreasing as long as \(r\) is small enough. How small \(r\) needs to be depends on \(T\), of course.

Also notice that \(B_t^* = b_t^*\) for all \(t \geq 1\) when \(r = 0\). Indeed, \(B_T^* = b_T^* = 1\), which implies that \(B_{T-1}^* = b_{T-1}^*\). Since \(B_{T-1}^* = 1\), then \(b_{T-1}^* = 1\) as well. Thus, \(B_{T-2}^* = b_{T-2}^*\) (given that \(b_T^* = b_{T-1}^* = 1\)), so that \(b_{T-2}^* = 1\) (as \(B_{T-2}^* = 1\)). A straightforward induction argument finishes the proof. Thus, explicit incentives are strictly decreasing in the limit case when \(r = 0\). By continuity, explicit incentives are strictly decreasing when \(r\) is small enough. In particular, we can have explicit incentives the highest when implicit career concern incentives are also the highest. This complementary is intuitive. If the worker is not too risk–averse, the equilibrium contract implies that in every period explicit incentives should account for most of total incentives, which should be at their highest level when the worker is young.

4 Discussion

We review here explanations proposed in the theoretical literature for the time pattern of (implicit and explicit) incentives and empirical evidence on the sensitivity of (executive or managerial) pay to performance. In light of the standard prediction that executive compensation should be negatively related to comparative or relative performance measures, as a means to insulate executives from aggregate uncertainty or other events they are unlikely to control (see recently Celentani and Loveira (2006) on this), we focus on evidence on the relationship between measures of absolute performance and compensation. For consistency with the theory (our model and most career concerns models in the literature have no prediction regarding the relationship between the strength of performance–based incentives and performance), we also abstract from evidence on the relationship between the performance sensitivity of current pay and future performance.\(^7\)

\(^7\)Note, however, that more senior executives are more likely to produce higher output in the presence of human capital accumulation. Then, performance–pay sensitivity can be correlated with subsequent perfor-
Theoretical Literature  A number of papers have investigated the relationship between the degree of uncertainty about individual talent and the strength of implicit incentives, on the one hand, and on the impact of time on explicit or implicit incentives under uncertainty, on the other. Recently, Kovrijnykh (2008) shows that, when ability is career–specific, wages can become more sensitive to a worker’s reputation in the labor market, since the market anticipates workers with good career matches to exert more effort and, thus, offers them a higher wage to compensate them for their higher ability and effort. For this reason, as in the standard model, in equilibrium workers who are less certain about their career prospects may exert more effort. Martinez (2009) reaches the opposite conclusion in a model of career concerns and job assignment. Specifically, he shows that effort may be more sensitive to changes in a worker’s reputation when priors are more concentrated. Thus, a change in a worker’s reputation may imply a greater change in the effort level firms expect him to exert. As a result, incentives to exert effort may be stronger in equilibrium if prior beliefs are more precise. Both papers, however, abstract from the consideration of explicit contracts. Ignoring, instead, possible incentive conflicts between firms and workers, Harris and Holmström (1982) analyze long–term wage contracts in an environment with symmetric uncertainty about ability under one–sided (firm) long–term commitment, and characterize their aggregate implications. In particular, they prove that equilibrium contracts prescribe wages that are downward rigid in a worker's age and increase only when the worker's market value increases above his current wage.

As for the relationship between explicit and implicit incentives, Baker, Gibbons, and Murphy (1994) analyze the optimal combination of subjective and objective performance measures in implicit and explicit incentive contracts, respectively. They identify circumstances under which objective and subjective measures are complements, that is, only an appropriate combination of the two yields a positive profit. More generally, they show that, if the objective measure becomes more accurate, then the optimal contract not only weighs more the objective measure of performance but also weighs more the subjective measure. Intuitively, as the improved objective measure increases the value of an employment rel-
tionship, it reduces a firm’s incentive to renege on rewards based on the subjective measure. Similarly, Levin (2003) examines optimal incentive provision, under either adverse selection or moral hazard, in a dynamic framework in which enforcement is imperfect and performance measures may be subjective. He identifies condition under which optimal self–enforcing contracts are stationary—the same compensation scheme is used at every date—and proves that under moral hazard optimal contracts entail termination under poor performance, when performance is subjectively assessed.

**Empirical Literature**  In the literature, the evidence on the importance of the relationship between the sensitivity of executive pay to performance and executive experience, and on its sign, is mixed. For instance, in a sample of 419 venture partnership agreements and offering memoranda for funds formed between 1978 and 1992, Gompers and Lerner (1999) find that compensation for older and larger venture capital organizations is *more sensitive* to performance than compensation for other venture groups. These differences are statistically significant whether the percentage of profits accruing to the venture capitalists (although results in this case are less conclusive) or the elasticity of compensation with respect to performance is examined. They also document that cross–sectional variation in compensation terms for younger, smaller venture organizations is considerably less than for older, larger organizations, that the fixed component of compensation is higher for smaller, younger funds, and funds focusing on high–technology or early–stage investments, and, finally, that no relationship can be detected between incentive compensation and performance. The authors interpret these finding as broadly consistent with the standard career concerns model.\(^8\)

On the contrary, using a longitudinal sample of 1,488 chief executive officers (CEO’s) followed from 1974–1985, Murphy (1986) finds that the sensitivity of pay to performance is *negatively* influenced by CEO experience (or years at the firm).\(^9\) Similarly, in a study of

\(^8\)Palia (2001) also analyzes the sensitivity of CEO pay to performance, taking into account managerial’s ownership of stock options in addition to stocks. His (fixed effects) estimates exhibit a positive relationship between CEO performance–pay sensitivity and CEO experience or age.

\(^9\)A prediction of his multi–period incentive model—in which current pay is assumed independent of current performance—is that the effect of performance on compensation increases with tenure, since rewards and penalties for performance are distributed over all remaining years of an employment contract. This channel could be a further explanation for the increase in the performance sensitivity of pay with experience or tenure.
CEOs of large commercial banks observed between 1982 and 1987, Barro and Barro (1990) estimate that the sensitivity of pay to performance diminishes with CEO experience. On the basis of longitudinal data on managers from a single company, Kahn and Sherer (1990) also find that variable compensation (bonuses) for managers who are in high-level positions, work at corporate headquarters, and have low seniority are more sensitive to performance than compensation paid to managers who do not satisfy these three criteria.\footnote{However, they report that merit pay seems to be granted on the same basis across managerial levels, plant locations, and seniority levels.}

Finally, a number of studies have cast doubt on the significance of the performance–pay relationship or reported evidence of substantial variation in its importance across firms of different size, industries, time periods, and countries. Among those, Baker, Jensen, and Murphy (1988) cite evidence on the lack of pay–performance sensitivity altogether, and in support of promotion–based incentive systems. Jensen and Murphy (1990) examine the link between changes in shareholder wealth and CEO pay, and find a significant positive relation between pay and performance, despite a decline in the pay–performance relation and the level of CEO pay since the 1930s. Murphy (1999) reports that total (cash compensation, stock, restricted stock, and stock options) pay–performance sensitivities vary with industry and company size, and documents that changes in these sensitivities from 1992 through 1996 were driven primarily by increases in stock–options incentives. More recently, in their review of the challenges associated with the determination of executive compensation, Bebchuk and Fried (2003) lament the documented lack of association between either incentive pay (bonus) or non–equity compensation (such as pensions, deferred pay, and loans) and performance, and they interpret it as the main reason why shareholders and regulators have increasingly considered equity–based compensation as a primary tool to link executive pay to performance. Finally, Fernandes, Ferreira, Matos, and Murphy (2009) document that the average CEO in the U.S. (as of 2006) receives 42\% of his pay in the form of options or stocks, more than twice the average percentage in other countries (20\%).\footnote{More recent figures for the U.S. are documented by the New York Times on April 3, 2010, based on the data on pay that the compensation research firm Equilar compiled for 200 chief executives at 199 public companies that filed their annual proxies by March 27 and had revenue of at least $6.3 billion—one company, Motorola, had co-C.E.O.’s. It is reported that the median pay package declined by 13 percent last year to $7.7 million, the average total pay decreased by 15 percent to $9.5 million, causing the median compensation}
Overall we interpret the literature as suggesting that: (i) a positive relationship exists between executive pay and (measures of firm’s) performance, albeit not too strong and variable in firm size, over time, and across industries and countries; and (ii) the strength of the sensitivity of pay to performance can either increase or decrease with executive experience. Note, however, that unobserved heterogeneity in executive compensation arrangements in place at a same or different firms may affect results derived from cross-sectional analyses of the relationship between CEO experience and the strength of performance pay. For instance, suppose abler executives are more likely to experience longer careers and, being their effort more productive, are also more likely to be rewarded through performance-based contracts highly contingent on firm outcomes. If so, the estimated relationship between the sensitivity of pay to performance and experience may just reflect the presence of such heterogeneity in any given sample of executives. In such case any estimate of the performance-pay relationship that does not take into account this heterogeneity is not interpretable as consistent (or inconsistent) with the career concerns model. Also, for the same reason endogeneity of contract choice, if boards or shareholders’ decisions on compensation are correlated with ability characteristics that they observe or that affect an executive’s effort, could potentially invalidate any estimate.

With these caveats in mind, we proceed to discuss briefly how we could test for the presence of learning-by-doing on the basis of available data and estimation strategies common in the literature. We also illustrate how the presence of human capital accumulation, if unaccounted for, would affect estimates of the strength and time profile of the sensitivity of pay to performance.

package back to the same level as in 2004. Among the C.E.O.’s of TARP (Troubled Assets Relief Program) recipients last year, 12 are on the Equilar list, including Kenneth D. Lewis of Bank of America, Vikram S. Pandit of Citigroup, John G. Stumpf of Wells Fargo, Lloyd C. Blankfein of Goldman Sachs, Jamie L. Dimon of JPMorgan Chase, and Kenneth I. Chenault of American Express. According to Equilar, the median pay for this group was $6 million in 2009, a 34 percent decrease from 2008. The median stock and options portion of their pay decreased by 94 percent and 92 percent, respectively. But the median cash payout for the group rose 20 percent. The highest-paid executive in the list, Lawrence J. Ellison of Oracle, received a compensation for 2009 of $84.5 million dollars, which consisted of a salary $1,000,000, a bonus of $3,586,813, ‘perks’ of $1,493,946, and options of $78,421,000. The least-paid executive (excluding Steven P. Jobs of Apple, who collected a $1 salary) was Mr. Lewis of Bank of America with $32,171 of cash compensation.
A Test Consider first the case in which the worker’s ability is not subject to shocks. In analogy with Gibbons and Murphy (1992) and assuming shareholders maximize their wealth, as measured by the market value of the firm’s common stock, suppose $y_t = V_t$, where $V_t$, as in Gibbons and Murphy, is interpreted as end-of-period firm value, so $\Delta V_t = y_t - E\{y_t|y_1,\ldots,y_{t-1}\}$ and

$$w_t = c_t + b_t y_t = c_t + b_t (\Delta V_t + E\{y_t|y_1,\ldots,y_{t-1}\}).$$

Recall that $c_t = (1-b_t) E\{y_t|y_1,\ldots,y_{t-1}\}$ and $E\{y_t|y_1,\ldots,y_{t-1}\} = m_{t-1} + \hat{a}_t + k(\hat{a}_1,\ldots,\hat{a}_{t-1})$. Thus,

$$w_t = (1-b_t) E\{y_t|y_1,\ldots,y_{t-1}\} + b_t (\Delta V_t + E\{y_t|y_1,\ldots,y_{t-1}\}) = E\{y_t|y_1,\ldots,y_{t-1}\} + b_t \Delta V_t = m_{t-1} + \hat{a}_t + k(\hat{a}_1,\ldots,\hat{a}_{t-1}) + b_t \Delta V_t.$$

Now, $\Delta V_{t-1} = y_{t-1} - m_{t-2} - \hat{a}_{t-1} - k(\hat{a}_1,\ldots,\hat{a}_{t-2})$. According to the interpretation of $y_t$ as end-of-period firm value, (17) in first differences ($\Delta w_t = w_t - w_{t-1}$) rewrites as

$$\Delta w_t = m_{t-1} + \hat{a}_t + k(\hat{a}_1,\ldots,\hat{a}_{t-1}) + b_t \Delta V_t - m_{t-2} - \hat{a}_{t-1} - k(\hat{a}_1,\ldots,\hat{a}_{t-2}) - b_{t-1} \Delta V_{t-1}$$

$$= \hat{a}_t - \hat{a}_{t-1} + k(\hat{a}_1,\ldots,\hat{a}_{t-1}) - k(\hat{a}_1,\ldots,\hat{a}_{t-2}) + b_t \Delta V_t + m_{t-1} - m_{t-2} - b_{t-1} \Delta V_{t-1}$$

$$= \Delta \hat{a}_t + \Delta k_t + b_t \Delta V_t + m_{t-1} - m_{t-2} - b_{t-1} \Delta V_{t-1}.$$

Let $\hat{k}_t = k(\hat{a}_1,\ldots,\hat{a}_{t-1})$. Observe that

$$m_{t-1} = \frac{\sigma^2_m + \sigma^2 \sum_{r=1}^{t-2}(y_r - \hat{a}_r - \hat{k}_r)}{\sigma^2 + (t-1)\sigma^2} = \frac{\sigma^2 m_1 + \sigma^2 \sum_{r=1}^{t-2}(y_r - \hat{a}_r - \hat{k}_r) + y_{t-1} - \hat{a}_{t-1} - \hat{k}_{t-1}}{\sigma^2 + (t-1)\sigma^2}$$

$$= \frac{\sigma^2 m_1 + \sigma^2 \sum_{r=1}^{t-2}(y_r - \hat{a}_r - \hat{k}_r)}{\sigma^2 + (t-1)\sigma^2} = \frac{\sigma^2 m_1 + \sigma^2 \sum_{r=1}^{t-2}(y_r - \hat{a}_r - \hat{k}_r) + y_{t-1} - \hat{a}_{t-1} - \hat{k}_{t-1}}{\sigma^2 + (t-1)\sigma^2}$$

$$= \frac{m_{t-2}[\sigma^2 + (t-2)\sigma^2]}{\sigma^2 + (t-1)\sigma^2} + \frac{\sigma^2 (y_{t-1} - \hat{a}_{t-1} - \hat{k}_{t-1})}{\sigma^2 + (t-1)\sigma^2} = \frac{m_{t-2}[\sigma^2 + (t-2)\sigma^2]}{\sigma^2 + (t-1)\sigma^2} + \frac{\sigma^2 (m_{t-2} + \Delta V_{t-1})}{\sigma^2 + (t-1)\sigma^2}$$

$$= \frac{m_{t-2}[\sigma^2 + (t-2)\sigma^2]}{\sigma^2 + (t-1)\sigma^2} + \frac{\sigma^2 \Delta V_{t-1}}{\sigma^2 + (t-1)\sigma^2} = \frac{\sigma^2 \Delta V_{t-1}}{\sigma^2 + (t-1)\sigma^2}$$

which implies

$$\Delta w_t = \Delta \hat{a}_t + \Delta \hat{k}_t + b_t \Delta V_t + \left[\frac{\sigma^2}{\sigma^2 + (t-1)\sigma^2} - b_{t-1}\right] \Delta V_{t-1} = \alpha_t + \phi(t) + \beta_t \Delta V_t + \gamma_t \Delta V_{t-1}$$

$$= \hat{a}_t + \beta_t \Delta V_t + \gamma_t \Delta V_{t-1}. \quad (18)$$
The first difference wage equation is a finite distributed–lag model with time/experience effects captured by $\alpha_t$, as in Gibbons and Murphy, with $\hat{\alpha}_t = \Delta \hat{\alpha}_t + \phi(t) = \alpha_t + \phi(t)$. Differently from Gibbons and Murphy, in addition to $\alpha_t$ our model admits a non–linear time effect. This extra term, $\Delta k_t = \hat{k}_t - \hat{k}_{t-1} = k(\hat{a}_1, \ldots, \hat{a}_{t-1}) - k(\hat{a}_1, \ldots, \hat{a}_{t-2}) = k(\hat{a}_{t-1}) = \phi(t)$, captures learning–by–doing on the job. Note that, given that $\hat{\alpha}_t$ is a deterministic function of time in equilibrium, the effort path can be approximated arbitrarily well by a suitable polynomial series $\phi(t)$. Moreover, $\phi(t)$ is an unobserved effect independent of ability if effort is constant across workers of different ability (conditional on tenure), as it is the case in the equilibrium of our model.

Observe that the model in first differences does not allow to formulate testable implications for the time path of the intercept $\alpha_t$ or $\phi(t)$. For this we would need a characterization of the speed of decline of effort in equilibrium, in addition to imposing a functional form assumption about the human capital accumulation function. However, the production function or output equation,

$$y_t = \hat{\alpha}_t + k(\hat{a}_1, \ldots, \hat{a}_{t-1}) + \theta + \varepsilon_t = \alpha_{Lt} + f(t) + \nu + \varepsilon_t$$

is interpretable as a partially linear static panel data model with strictly exogenous covariates, with $\alpha_{Lt} = \hat{\alpha}_t$, $f(t) = k(\hat{a}_1, \ldots, \hat{a}_{t-1})$, and $\nu = \theta$. Given the presence of the unobserved individual effect, $\nu$, the model can be readily estimated as a random effects model (and inference conducted based on robust covariance estimation, to allow for potential time variation and auto–correlation in the error term). Therefore, a simple $t$– or $F$–test on $f(t)$, if a parametric form for $f(t)$ is specified—testing is also possible in the semi–parametric case (see Härdle, Liang, and Gao (2000))—could allow to detect the presence of human capital acquisition. Clearly, a testable implication from our model is that both the output process and the wage difference process should display an upward time/experience trend.

A few observations may help clarify our proposed exercise. First, note that in most specifications (for instance, in Gibbons and Murphy) the intercept of the wage equation in first differences is allowed to vary over time, where time is intended as an executive’s experience, tenure at the firm, or years left in office. However, such a specification is not immune to potential misspecification problems, if the human capital accumulation process is highly
nonlinear in time (see equation (34) in Gibbons and Murphy). Second, the uncorrelation between the unobserved effect and the included explanatory variables, which motivated our interpretation of the output equation as a random effect model, is an implication of equilibrium in our framework and in most career concerns models. Since the optimal effort choice should not vary in $\theta$, the unobserved effect $\nu$ should be uncorrelated with $\alpha_{Lt}$ and $f(t)$—for this, in our model both the linear separability between $\theta_0$ and $k_t$ and the independence of $\theta_0$ from $\hat{a}_t$ are crucial. Therefore, a test on the validity of the random effect assumption is also an indirect test of our theory. Third, suppose, instead, we interpret the output equation as a fixed effect model. In light of our assumption of linear separability of the human capital investment over time, we have

$$\Delta y_t = \Delta \hat{a}_t + \Delta \hat{k}_t + \Delta \epsilon_t = \Delta \hat{a}_t + k(a_{t-1}) + \Delta \epsilon_t.$$ 

Including a polynomial approximation term for $k(a_{t-1})$, $\phi(t)$, in the first difference output equation could allow to test for the presence of human capital accumulation in a much more general framework, in which other unobserved (time–invariant) terms, in addition to $\nu$, are correlated with included variables and which is robust to the omission of any time–invariant regressors.\textsuperscript{12}

As mentioned, analogously to the standard omitted variable problem, if human capital acquisition is present and not explicitly controlled for, estimation of the effects of interest may be problematic. We turn now to illustrate this point in further detail. For clarity, let now $t$ denote calendar time and $\tau_t$ an executive’s experience, tenure at the firm, or years in office in period $t$. Suppose $\phi(\tau_t)$ can be accurately specified as a third–degree polynomial in time. We can rewrite

$$\Delta w_t = \alpha_t + \phi(\tau_t) + \beta_t \Delta V_t + \gamma_t \Delta V_{t-1} = \alpha_t + \zeta_1 \tau_t + \zeta_2 \tau_t^2 + \zeta_3 \tau_t^3 + \beta_t \Delta V_t + \gamma_t \Delta V_{t-1}$$

$$= \alpha_t + [\Delta V_t, \Delta V_{t-1}] [\beta_t, \gamma_t] + [\tau_t, \tau_t^2, \tau_t^3] [\zeta_1, \zeta_2, \zeta_3] = \alpha_t + X_{1t} \delta_{1t} + X_{2t} \delta_2.$$ 

\textsuperscript{12}Admittedly, our formulation of the problem does not allow for any variation in the performance–pay sensitivity parameter across firms of different size, industries, or countries for executives with the same experience (or tenure or years in office, depending on the relevant time dimension for human capital accumulation). Ideally, based on sufficient sample size and compatibly with the observed turnover pattern of executives, the output and wage equation should be estimated on different sub–samples.
For the moment, ignore for simplicity intercept year dummies and the interaction terms between year dummies, on the one hand, and the current and lagged differenced performance measures $\Delta V_t$ and $\Delta V_{t-1}$, on the other. Then, the above equation becomes $\Delta w_t = \alpha + X_t \delta_1 + X_2 \delta_2$. Estimating this equation by least squares omitting $X_2$ (a matrix of dimension $NT \times 3$), and after appending the error $\varepsilon_t$, we obtain

$$
E(\hat{\delta}_1) = \delta_1 + (X'_1 X_1)^{-1} X'_1 X_2 \delta_2
$$

where, as standard, each column of the $2 \times 3$ matrix $(X'_1 X_1)^{-1} X'_1 X_2$ is the column of slopes in the least square regression of the corresponding column of $X_2$ on the columns of $X_1$. Let $i$ denote an individual executive. For instance, the first column in $(X'_1 X_1)^{-1} X'_1 X_2$ is the column of the two slopes of the multivariate regression of $\tau$ ($NT \times 1$ column vector) on $\Delta V$ ($NT \times 1$ column vector) and $\Delta V_{-1}$ ($NT \times 1$ column vector), respectively. Denote these coefficients by $b_{11}$ and $b_{21}$, respectively. Then,

$$
b_{11} = \frac{\sum_{it}(\tau_{it} \Delta V_{it}) \sum_{it}(\Delta V_{it})^2 - \sum_{it}(\Delta V_{it} \Delta V_{it-1}) \sum_{it}(\tau_{it} \Delta V_{it})}{\sum_{it}(\Delta V_{it})^2 \sum_{it}(\Delta V_{it-1})^2 - [\sum_{it}(\Delta V_{it} \Delta V_{it-1})]^2}
$$

and

$$
b_{21} = \frac{\sum_{it}(\tau_{it} \Delta V_{it-1}) \sum_{it}(\Delta V_{it})^2 - \sum_{it}(\Delta V_{it} \Delta V_{it-1}) \sum_{it}(\tau_{it} \Delta V_{it})}{\sum_{it}(\Delta V_{it})^2 \sum_{it}(\Delta V_{it-1})^2 - [\sum_{it}(\Delta V_{it} \Delta V_{it-1})]^2}.
$$

The remaining coefficients can be obtained analogously. Allowing $\alpha_t$, $\beta_t$ and $\gamma_t$ to vary over time implies that the intercept and slope parameters for the differenced values are merely estimated on the cross section. In this latter case, depending on the importance of positive and negative terms in the expressions for $b_{11}$ and $b_{21}$—in particular on the size of the covariance between $\Delta V_{it}$ and $\Delta V_{it-1}$—for $\sum_{i}(\Delta V_{it} \Delta V_{it-1})$ large (not unlikely) and increasing over time it is possible that $\hat{\delta}_it$ will be inflated in later periods and attenuated in early periods. Then, controlling for human capital accumulation in the specification of the first–difference equation for pay could lead to opposite conclusions on the time pattern of incentives than the ones Gibbons and Murphy reached.\(^{13}\) This, in turn, could explain the

\(^{13}\)Even in Gibbons and Murphy (1992), the evidence on the increasing power of explicit performance–based pay as an executive nears retirement is partly mixed. In the non–logarithmic version of the main regression equation that relates changes in executives’ compensation to changes in shareholders’ wealth, the coefficient on the interaction term ‘shareholder’s wealth x few years left’ is positive but insignificant, suggesting that
mixed evidence on the time/tenure/experience/age profile of the sensitivity of executive pay to performance.

Finally, consider the case in which ability is subject to shocks and note that

\[ y_t = \hat{a}_t + k_t + \theta_0 + \sum_{s=1}^{\tau_t-1} \epsilon_s + \epsilon_t = \alpha L_t + f(\tau_t) + \nu + \epsilon_t + \sum_{s=1}^{\tau_t-1} \epsilon_s = \alpha L_t + f(\tau_t) + \nu + \hat{\epsilon}_t \]

where \( \epsilon_t \)'s are independently and identically normally distributed. The difference with respect to the previous case, in which human capital is not subject to shocks, is that now the error includes a moving-average component from the cumulative effect of the shocks to ability, with \( \text{Var}(\hat{\epsilon}_t) = \text{Var}(\epsilon_t + \sum_{s=1}^{\tau_t-1} \epsilon_s) = \sigma^2 + (\tau_t - 1)\sigma^2. \) In this case statistical inference on the parameters requires robust covariance matrix estimation. However, as the asymptotic analysis (based on fixed \( T \) and large \( N \)) is unaltered and, again, in light of the conditional independence of \( \hat{a}_t \) and \( \hat{k}_t \) from \( \theta_0 \), standard random effects estimation is appropriate. Hence, a testing procedure to detect the presence of human capital accumulation, similar to the above, obtains. The same argument as above applies to the estimation of the performance–pay sensitivity of compensation.

5 Conclusion

In this paper we introduce human capital accumulation in a life–cycle model of career concerns and analyze its implications for optimal incentive provision. We show that the tenure effect and the possibility for a worker to affect changes in his ability have a substantial impact on the relationship between implicit and explicit incentives over the working life cycle. Specifically, we show that well–known results on the evolution of effort with experience and on the substitutability between implicit and explicit incentives can be reversed. First, implicit incentives for effort disappear over time even when uncertainty about talent is exogenously replenished. Second, it can be optimal for explicit incentives to complement rather than substitute implicit incentives. Overall our results help explain the conflicting evidence on the pay–performance relation is independent of years remaining as CEO. Instead, the relation in log–terms is positive and significant. The authors favor this latter formulation and motivate it as controlling for the impact of firm size on the pay–performance relation. In fact, they find that the pay–performance relation varies significantly with firm size but that the pay–performance elasticity is almost invariant to firm size.
the sensitivity of pay to performance for managers at different stages of their working life.

References


6 Appendix

6.1 Derivation of Equation (5)

Note that (5) reduces to (3) when \( s = t + 1 \). Suppose now, by induction, that (5) holds for some \( s \geq t + 1 \). We are done if we show that (5) also holds for \( s + 1 \). For this, notice, by (3), that

\[
m_{s+1}(y^t, y_t, \ldots, y_s) = \mu_s m_s(y^t, y_1, \ldots, y_{s-1}) + (1 - \mu_s) \left[ y_s - \hat{a}_s - \sum_{r=1}^{s-1} k(\hat{a}_r) \right].
\]

Hence, by the induction hypothesis (and recalling that \( \prod_{r=s+1}^t \mu_r = 1 \)),

\[
m_{s+1}(y^t, y_t, \ldots, y_s) = \mu_s m_s(y^t) \prod_{r=t}^{s} \mu_r + \mu_s \sum_{q=t}^{s-1} (1 - \mu_q) \prod_{r=q+1}^{s-1} \mu_r \left[ y_q - \hat{a}_q - \sum_{r=1}^{q-1} k(\hat{a}_r) \right]
\]

\[+ (1 - \mu_s) \prod_{r=s+1}^{s-1} \mu_r \left[ y_s - \hat{a}_s - \sum_{r=1}^{s-1} k(\hat{a}_r) \right]
\]

\[= m_t(y^t) \prod_{r=t}^{s} \mu_r + \sum_{q=t}^{s} (1 - \mu_q) \prod_{r=q+1}^{s} \mu_r \left[ y_q - \hat{a}_q - \sum_{r=1}^{q-1} k(\hat{a}_r) \right],
\]

which is the desired result.

6.2 Proof of Proposition 1

Here we show that the only equilibria in the benchmark case are uncontingent equilibria. For this, suppose the market expects the worker to follow a strategy \( \hat{\sigma} \) that is uncontingent from period \( t + 1 \) on, with \( t \leq T - 1 \), and let \( \hat{a}_s \) be the choice of effort in period \( s \geq t + 1 \) prescribed by \( \hat{\sigma} \). The worker’s wage in period \( s \geq t + 1 \) is then given by

\[
w_s(y^t, y_t, \ldots, y_{s-1}) = m_s(y^t, y_t, \ldots, y_{s-1}) + \hat{a}_s + k_t(y^t) + k(\hat{\sigma}_t(y^t)) + \sum_{q=t+1}^{s-1} k(\hat{a}_q);
\]

\( k_t(y^t) \) is the worker’s human capital in period \( t \) if he follows \( \hat{\sigma} \) and his output history is \( y^t \). An argument similar to the one used to derive (5) from (3) shows that for all \( s \geq t + 1 \),

\[
m_s(y^t, y_t, \ldots, y_{s-1}) = m_t(y^t) \prod_{r=t}^{s-1} \mu_r + (1 - \mu_t) \prod_{r=t+1}^{s-1} \mu_r [y_t - \hat{\sigma}_t(y^t) - k_t(y^t)]
\]

\[+ \sum_{q=t+1}^{s-1} (1 - \mu_q) \prod_{r=q+1}^{s-1} \mu_r \left[ y_q - \hat{a}_q - k_t(y^t) - k(\hat{\sigma}_t(y^t)) - \sum_{r=t+1}^{q-1} k(\hat{a}_r) \right].
\]

35
Now assume the worker behaves according to \( \tilde{\sigma} \) from period \( t + 1 \) on and let \( V_t(a|z^t) \) be his (lifetime) payoff from choosing \( a \) in period \( t \) when his history is \( z^t = (a^t, y^t) \). Since

\[
E[y_t - \tilde{\sigma}_t(y^t) - k_t(y^t)|a, z^t] = E[\theta_t|z^t] + a - \tilde{\sigma}_t(y^t)
\]

and

\[
E[y_q - \hat{a}_q - k_t(y^t) - k(\tilde{\sigma}_t(y^t)) - \sum_{r=t+1}^{q-1} k(\hat{a}_r)|a, z^t] = E[\theta_q|z^t] + k(a) - k(\tilde{\sigma}_t(y^t)),
\]

for all \( q \geq t + 1 \), we then have that

\[
V_t(a|z^t) = -E\left[\exp\left(-r\left\{\sum_{s=t+1}^{T} \delta^{s-t} \left[w_s - g(a_s)\right]\right\}\right)\right]
\]

\[
= -C \exp\left(-r\left\{\sum_{s=t+1}^{T} \delta^{s-t} \left[1 - \mu_t \prod_{\tau=t+1}^{s-1} \mu_{+} + \sum_{q=t+1}^{s-1} (1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_{+} k(a)\right] - g(a)\right\}\right),
\]

where \( C \) is a term that is independent of \( a \); recall that if \( x \sim N(m, \sigma^2) \), then \( E[\exp(-rx)] = \exp(-rm + r^2\sigma^2/2) \). Since \( V_t(a|z^t) \) is strictly quasi-concave in \( a \), the necessary and sufficient first-order condition for the optimal choice of \( a \) in period \( t \) is

\[
\sum_{s=t+1}^{T} \delta^{s-t} \left[1 - \mu_t \prod_{\tau=t+1}^{s-1} \mu_{+} + \sum_{q=t+1}^{s-1} (1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_{+} k'(a)\right] = g'(a),
\]

which is equation (7) in the main text.

We know that the solution to the above first-order condition is independent of \( z^t \) and the market’s conjecture about the worker’s behavior. Hence, if the worker’s choice of effort from \( t + 1 \) on is uncontingent, his choice of effort in period \( t \) is also uncontingent. Since in any equilibrium the worker’s choice of effort in period \( T \) is always zero, and thus uncontingent, we can then conclude, by backward induction, that every equilibrium is uncontingent. This establishes the desired result.

### 6.3 Proof of Proposition 5

We proceed by backward induction.

**Last Period.** Let \( (\sigma^*, \omega^*) \) be a candidate equilibrium and consider the worker in period \( T \). Suppose the worker’s history is \( z^T = (a^T, y^T) \) and let \( m_T = m_T(z^T) \) and \( k_T = k_T(z^T) \) be the
mean of the worker’s belief about his ability and the worker’s human capital, respectively. If the worker’s choice of effort is \( a \), then \( y_T = a + k_T + \theta_T + \varepsilon_T \) is normally distributed with mean \( a + k_T + m_T \) and variance \( \sigma_T^2 \). If \( w_T = c_T + b_T y_T \), the worker’s expected lifetime payoff from choosing \( a \) in period \( T \) conditional on the history \( z^T \) is
\[
V_T(a|z^T) \propto -\mathbb{E} \left[ \exp \left( -r \{ w_T - g(a) \} \right) \right] = -C \exp \left( -r \{ c_T + b_T [m_T + k_T + a] - g(a) \} \right),
\]
where \( C \) is a term independent of \( a \); recall that if \( x \sim \mathcal{N}(m, \sigma^2) \), then \( \mathbb{E} \left[ \exp(-rx) \right] = \exp(-rm + r^2\sigma^2/2) \). Since \( V_T(a|z^T) \) is strictly quasi-concave in \( a \), a necessary and sufficient condition for the optimal choice of \( a \) is that
\[
g'(a) = b_T. \tag{19}
\]

Denote the solution to (19) by \( a_T^* = a_T^*(b_T) \). Notice that \( a_T^* \) does not depend on \( z^T \). Thus, in any equilibrium, the worker’s choice of effort in period \( T \) is uncontingent.

Let \( \hat{m}_T = \hat{m}_T(y^T) \) and \( \hat{k}_T = \hat{k}_T(y^T) \) be the market’s conjectures about the worker’s mean ability and human capital, respectively. Both conjectures can only depend on the observable part of the worker’s history and must be correct in equilibrium. Now let
\[
c_T(b_T) = (1 - b_T)[\hat{m}_T(y^T) + \hat{k}_T(y^T) + a_T^*(b_T)], \]
\[
X_T(b_T) = c_T(b_T) + b_T[a_T^*(b_T) + \hat{k}_T + \theta_T + \varepsilon_T] - g(a_T^*(b_T)),
\]
and suppose the worker behaves according to \( \sigma^* \) up to period \( T \). Notice that both \( c_T(b_T) \) and \( X_T(b_T) \) depend on \( y^T \). By construction, \( U_T(b_T|\sigma^*, y^T) = -\mathbb{E}[\exp(-rX_T(b_T))]. \) Since \( X_T(b_T) \) is normally distributed with mean \( \hat{m}_T + \hat{k}_T + a_T^*(b_T) - g(a_T^*(b_T)) \) and variance \( \Sigma_T^2 = \sigma^2_T + \sigma^2_T \), we then have that
\[
U_T(b_T|\sigma^*, y^T) = -\exp \left( -r \left\{ \hat{m}_T(y^T) + \hat{k}_T(y^T) + a_T^*(b_T) - g(a_T^*(b_T)) - \frac{1}{2} r^2 b_T^2 \Sigma_T^2 \right\} \right). \]

The values of \( b_T \) that maximize \( U_T(b_T|\sigma^*, y^T) = U_T(b_T|y^T) \) are the possible equilibrium piece rates in period \( T \). It is immediate to see that the values of \( b_T \) that do so are the same for all \( y^T \in Y^T \). Now let \( h_T(b_T) = a_T^*(b_T) - g(a_T^*(b_T)) \) and notice, by (19), that
\[
h'_T(b_T) = (1 - b_T) \frac{\partial a_T^*(b_T)}{\partial b_T} = \frac{1 - b_T}{g''(a_T^*(b_T))}. \tag{20}
\]
Since \( g''' \geq 0 \), the function \( h_T \) is concave. Therefore, \( U_T(b_T|y^T) \) is strictly quasi–concave in \( b_T \), and so there exists a unique value \( b^*_T \) of \( b_T \) that maximizes \( U_T(b_T|y^T) \) regardless of \( y^T \). We obtain \( b^*_T \) by solving the (necessary and sufficient) first–order condition for \( \max_{b_T} U_T(b_T|y^T) \).

From (20), this condition is

\[
H_T(b_T) = 1 - b_T - rb_T\Sigma_T^2 g''(a^*_T(b_T)) = 0. \tag{21}
\]

Notice that \( b^*_T \), which is positive, does not depend on the market’s conjecture about the worker’s behavior.

To summarize, we have established that every equilibrium is uncontingent in period \( T \). Moreover, the worker’s effort \( a^*_T \) in period \( T \) and the piece rate \( b^*_T \) in period \( T \) are the unique solutions to \( g'(a_T) = b^*_T \) and \( H_T(b_T) = 0 \), respectively.

**Induction Step.** Let \((\sigma^*, \omega^*)\) be a candidate equilibrium and suppose, by induction, that there exists a period \( t \leq T - 1 \) for which (i) to (iii) below hold:

(i) The pair \((\sigma^*, \omega^*)\) is uncontingent from period \( t + 1 \) on;

(ii) Let \( b^*_s \) is the piece rate in period \( s \geq t + 1 \) and, for all \( q \geq t + 1 \), let

\[
B^*_q = b^*_q + \sum_{s=q+1}^{T} \delta^{s-q}(1 - b^*_s)(1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_\tau.
\]

Now let \( a^*_s(B) \) denotes the unique solution to

\[
g'(a) = B + \sum_{\tau=s+1}^{T} \delta^{\tau-s}B^*_\tau k'(a).
\]

The worker’s worker in period \( s \geq t + 1 \) is \( a^*_s = a^*_s(B^*_s) \), where \( B^*_s \) is the unique solution to

\[
H_s(B) = 1 - B - r \left[ g''(a^*_s(B)) - k''(a^*_s(B)) \sum_{\tau=s+1}^{T} \delta^{\tau-s}B^*_\tau \right] \left[ \Sigma_s^2B + \sigma_s^2 \sum_{\tau=s+1}^{T} \delta^{\tau-s}B^*_\tau \right] + k'(a^*_s(B)) \sum_{\tau=s+1}^{T} \delta^{\tau-s}(1 - B^*_\tau) = 0.
\]

(iii) \( \sum_{\tau=t+1}^{T} \delta^{\tau-t}B^*_\tau \geq 0 \) if \( k \) is non–linear.

Note that (i) to (iii) are satisfied in period \( T - 1 \). We want to show that (i) to (iii) hold in period \( t - 1 \) as well.
Consider the worker in period $t$ and suppose his history is $z^t = (a^t, y^t)$. Let $m_t = m_t(z^t)$ be the mean of the the worker’s belief about his ability and $k_t = k_t(z^t)$ be the worker’s human capital. Moreover, let $\widehat{a}_s, \widehat{m}_s = \widehat{m}_s(y^s)$, and $\widehat{k}_s = \widehat{k}_s(y^s)$ be the market’s conjectures about the worker’s choice of effort, expected ability, and human capital in period $s \geq t$, respectively. These conjectures must be correct in equilibrium. In particular, $\widehat{a}_s = a^*_s$ for all $s \geq t + 1$. Also notice that in order for $(\sigma^*, \omega^*)$ to be an equilibrium, it must be that $c_s = c^*_s$ for all $s \geq t + 1$, where

$$c^*_s = (1 - b^*_s)[\widehat{m}_s + \widehat{a}_s + \widehat{k}_s].$$

Suppose the worker behaves according to $\sigma^*$ from period $t + 1$ on, and let $a$ be his choice of effort in period $t$. Moreover, suppose the linear contract offered to the worker in period $t$ is $(b_t, c_t)$ and let

$$B_t = b_t + \sum_{s=t+1}^{T} \delta^{s-t}(1 - b^*_s)(1 - \mu_t) \prod_{\tau=t+1}^{s-1} \mu_{\tau}. $$

We have the following result

**Lemma 1.** The random variable $c_t + b_t y_t - g(a) + \sum_{s=t+1}^{T} \delta^{s-t}[c^*_s + b^*_s y_s - g(a^*_s)]$ is normally distributed with mean

$$B_t a + \sum_{q=t+1}^{T} \delta^{q-t} B^*_q h(a) - g(a) + D,$$

where $D$ is a term that is independent of $a$.

**Proof:** We begin with the following auxiliary result.

**Lemmata:** For all $t \leq T - 1$,

$$\sum_{s=t+1}^{T} \delta^{s-t}(1 - b_s) \sum_{q=t}^{s-1} \left( \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right) \xi_q = \sum_{q=t}^{T} \left[ \sum_{s=q+1}^{T} \delta^{s-t}(1 - b_s) \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right] \xi_q. \tag{22}$$

**Proof:** First notice that (22) holds if $t = T - 1$. Suppose now, by induction, that (22) holds
for some \( t \leq T - 1 \) and notice that

\[
\sum_{s=t}^{T} \delta^{s-t+1}(1 - b_s) \sum_{q=t}^{s-1} \left( \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right) \xi_q
\]

\[
= \delta(1 - b_t)\xi_{t-1} + \sum_{s=t+1}^{T} \delta^{s-t+1}(1 - b_s) \sum_{q=t}^{s-1} \left( \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right) \xi_q
\]

\[
= \delta(1 - b_t)\xi_{t-1} + \left[ \sum_{s=t}^{T} \delta^{s-t+1}(1 - b_s) \prod_{\tau=t}^{s-1} \mu_{\tau} \right] \xi_{t-1} + \sum_{s=t+1}^{T} \delta^{s-t+1}(1 - b_s) \sum_{q=t}^{s-1} \left( \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right) \xi_q
\]

\[
= \left[ \sum_{s=t}^{T} \delta^{s-t+1}(1 - b_s) \prod_{\tau=t}^{s-1} \mu_{\tau} \right] \xi_{t-1} + \sum_{s=t+1}^{T} \delta^{s-t+1}(1 - b_s) \sum_{q=t}^{s-1} \left( \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right) \xi_q.
\]

Hence, by the induction hypothesis,

\[
\sum_{s=t}^{T} \delta^{s-t+1}(1 - b_s) \sum_{q=t}^{s-1} \left( \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right) \xi_q
\]

\[
= \left[ \sum_{s=t}^{T} \delta^{s-t+1}(1 - b_s) \prod_{\tau=t}^{s-1} \mu_{\tau} \right] \xi_{t-1} + \sum_{q=t}^{T} \left[ \sum_{s=q+1}^{T} \delta^{s-t+1}(1 - b_s) \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right] \xi_q
\]

\[
= \sum_{q=t-1}^{T} \left[ \sum_{s=q+1}^{T} \delta^{s-t+1}(1 - b_s) \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right] \xi_q,
\]

and so (22) holds for \( t - 1 \) as well. The desired result now follows. 

Now observe that \( y_t = a_t + k_t + \theta_t + \varepsilon_t \), where \( k_t = k_t(z^t) \) is the worker’s human capital in period \( t \), and

\[
y_s = a^*_s + k_t + k(a) + \sum_{q=t+1}^{s-1} k(a^*_q) + \theta_s + \varepsilon_s
\]

for all \( s \geq t + 1 \). Hence,

\[
b_t y_t + \sum_{s=t+1}^{T} \delta^{s-t} b_s^* y_s = b_t a + \sum_{s=t+1}^{T} \delta^{s-t} b_s^* k(a) + D_1,
\]

(23)
where $D_1$ is a term independent of $a$. Moreover, by (22),

$$
\sum_{s=t+1}^{T} \delta^{s-t}(1 - b^*_s) \sum_{q=t}^{s-1} \left( \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right) (1 - \mu_q) [y_q - \hat{a}_q - \hat{k}_q]
$$

$$
= \sum_{q=t}^{T} \sum_{s=q+1}^{T} \delta^{s-t}(1 - b^*_s) \prod_{\tau=q+1}^{s-1} \mu_{\tau} \left( 1 - \mu_t \right) [y_t - \hat{a}_t - \hat{k}_t]
$$

$$
= \sum_{s=t+1}^{T} \delta^{s-t}(1 - b^*_s) \left( \prod_{\tau=t+1}^{s-1} \mu_{\tau} \right) \left( 1 - \mu_t \right) \left( 1 - \mu_t \right) a
$$

$$
+ \sum_{q=t}^{T} \delta^{q-t} \left[ \sum_{s=q+1}^{T} \delta^{s-q} (1 - b^*_s) \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right] (1 - \mu_q) k(a) + D_2,
$$

(24)

where $D_2$ is a term independent of $a$. Since (5) implies that

$$
\hat{m}_s = \left( \prod_{\tau=t}^{s-1} \mu_{\tau} \right) \hat{m}_t + \sum_{q=t}^{s-1} \left( \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right) (1 - \mu_q) [y_q - \hat{a}_q - \hat{k}_q]
$$

for all $s \geq t + 1$, equations (23) and (24) then imply that

$$
c_t + b_t y_t + \sum_{s=t+1}^{T} \delta^{s-t} [(1 - b^*_s)\hat{m}_s + b^*_s y_s]
$$

$$
= \left[ b_t + \sum_{s=t+1}^{T} \delta^{s-t} (1 - b^*_s) (1 - \mu_t) \prod_{\tau=s+1}^{r-1} \mu_{\tau} \right] a
$$

$$
+ \sum_{q=t}^{T} \delta^{q-t} \left\{ b^*_q + \sum_{s=q+1}^{T} \delta^{s-q} (1 - b^*_s) (1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_{\tau} \right\} k(a) + D_3,
$$

where $D_3$ is a term that is independent of $a$. Lemma 1 follows from this last equation and the fact that $c^*_s = (1 - b^*_s)[\hat{m}_s + \hat{a}_s + \hat{k}_s]$.

Now let $V_t(a|z^t)$ be worker’s expected lifetime payoff from choosing $a$ in period $t$ conditional on the history $z^t$. Lemma 1 implies that

$$
V_t(a|z^t) \propto -\mathbb{E} \exp \left\{ -r \left\{ c_t + b_t y_t - g(a) + \sum_{s=t+1}^{T} \delta^{s-t} [c^*_s + b^*_s y_s - g(a^*_s)] \right\} \right\}
$$

$$
= -C \exp \left( -r \left\{ B_t a + \sum_{q=t+1}^{T} \delta^{q-t} B^*_q k(a) - g(a) \right\} \right),
$$
where $C$ is a term that does not depend on $a$. Since $V_t(a|z^t)$ is strictly quasi-concave in $a$, a necessary and sufficient condition for the optimal choice of $a$ is that

$$g'(a) = B_t + \sum_{q=t+1}^T \delta^{q-t} B_q^* k'(a).$$

(25)

Denote the solution to (25) by $a^*_t(b_t, b_{t+1}^*, \ldots, b_T^*) = a^*_t(b_t)$. Notice that there is a one-to-one map between $b_t$ and $B_t$ and that $a^*_t(b_t)$ does not depend on $z^t$. Therefore, in order for $(\sigma^*, \omega^*)$ to be an equilibrium, it must be that the worker’s behavior in period $t$ is uncontingent.

We now show that the piece rate in period $t$ must also be uncontingent if $(\sigma^*, \omega^*)$ is to be an equilibrium. For this,

$$y_t(b_t) = a^*_t(b_t) + \hat{k}_t + \theta_t + \epsilon_t,$$
$$c_t(b_t) = (1 - b_t)E[y_t(b_t)],$$

and, for each $s \geq t + 1$, let

$$y_s(b_t) = \hat{a}_s + \hat{k}_s(b_t) + \theta_s + \epsilon_s,$$
$$c_s(b_t) = (1 - b^*_s)E[y_s(b_t)\mid y_t, \ldots, y_{s-1}],$$

where $\hat{k}_s(b_t) = \hat{k}_t + k(a^*_t(b_t)) + \sum_{q=t+1}^{s-1} k(\hat{a}_q)$. Moreover, let

$$X_t(b_t) = c_t(b_t) + b_t y_t(b_t) - g(a^*_t(b_t)) + \sum_{s=t+1}^T \delta^{s-t} [c_s(b_t) + b^*_s y_s(b_t) - g(\hat{a}_s)]$$

and suppose the worker behaves according to $\sigma^*$ up to period $t$; note that $X_t$ depends on $y^t$.

By construction, $U_t(b_t|\sigma^*, y^t) = -E[\exp(-rX_t(b_t))]$. We have the following result.

**Lemma 2.** $X_t(b_t)$ is normally distributed with mean $M_t$ and variance $V_t$, where

$$M_t = a^*_t(b_t) - g(a^*_t(b_t)) + \sum_{s=t+1}^T \delta^{s-t} k(a^*_t(b_t)) + D,$$
$$V_t = \left( B_t + \sum_{s=t+1}^T \delta^{s-t} B_s^* \right)^2 \sigma^2 + B_t^2 \sigma^2 + \sum_{s=t+1}^T (\delta^{s-t} B_s^*)^2 [\sigma^2 + (s - t)\sigma^2],$$

and $D$ is a constant that is independent of $a^*_t(b_t)$.
Proof: It is immediate to see that \( X_t(b_t) \) is normally distributed with mean \( M_t \). In order to prove that \( X_t(b_t) \) has variance \( V_t \), notice first that

\[
\begin{align*}
\sum_{s=t+1}^{T} \sigma_{s-t} b_s^*(b_t) &= \sum_{s=t+1}^{T} \delta^{s-t} b_s^*(b_t) = b_t(\theta_t + \varepsilon_t) + \sum_{s=t+1}^{T} \delta^{s-t} b_s^*(\theta_s + \varepsilon_s) + D_1, \\
\end{align*}
\]

where \( D_1 \) is a constant term. Also notice that for all \( s \geq t + 1 \),

\[
c_s^*(b_t) = (1 - b_s^*)[\widetilde{m}_s(b_t) + \widetilde{a}_s + \widetilde{k}_s(b_t)],
\]

where

\[
\widetilde{m}_s = (\prod_{\tau=t}^{s-1} \mu_\tau) \mu_t + \sum_{q=t}^{s-1} \left( \prod_{\tau=q+1}^{s-1} \mu_\tau \right) (1 - \mu_q)[y_q(b_t) - \hat{a}_q - \hat{k}_q(b_t)].
\]

Since, by the lemmata in the proof of Lemma 1,

\[
\sum_{s=t+1}^{T} \delta^{s-t}(1 - b_s^*) \sum_{q=t}^{s-1} \left( \prod_{\tau=q+1}^{s-1} \mu_\tau \right) (1 - \mu_q)(\theta_q + \varepsilon_q)
\]

\[
= \sum_{q=t}^{T} \delta^{q-t} \left[ \sum_{q+1}^{T} \delta^{q-q}(1 - b_s^*)(1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_\tau \right] (\theta_t + \varepsilon_t)
\]

\[
= \left[ \sum_{s=t+1}^{T} \delta^{s-t}(1 - b_s^*)(1 - \mu_t) \prod_{\tau=t+1}^{s-1} \mu_\tau \right] (\theta_t + \varepsilon_t)
\]

\[
+ \sum_{q=t+1}^{T} \delta^{q-t} \left[ \sum_{s=q+1}^{T} \delta^{s-q}(1 - b_s^*)(1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_\tau \right] (\theta_q + \varepsilon_q),
\]

we then have that

\[
X_t(b_t) = \left[ \sum_{s=t+1}^{T} \delta^{s-t}(1 - b_s^*)(1 - \mu_t) \prod_{\tau=t+1}^{s-1} \mu_\tau \right] (\theta_t + \varepsilon_t)
\]

\[
+ \sum_{q=t+1}^{T} \delta^{q-t} \left[ b_q^* + \sum_{s=q+1}^{T} \delta^{s-q}(1 - b_s^*)(1 - \mu_q) \prod_{\tau=q+1}^{s-1} \mu_\tau \right] (\theta_q + \varepsilon_q) + D_2,
\]

where \( D_2 \) is a constant term. Now notice that \( \theta_q = \theta_t + \sum_{s=t}^{q-1} \eta_s \) for all \( q \geq t + 1 \), and so

\[
X_t(b_t) = \left( B_t + \sum_{q=t+1}^{T} \delta^{q-t} B_q^* \right) \theta_t + B_t \varepsilon_t + \sum_{q=t+1}^{T} \delta^{q-t} B_q^* \left( \varepsilon_q + \sum_{s=t}^{q-1} \eta_s \right) + D_2,
\]

from which the desired result follows. \( \square \)
The possible equilibrium piece rates in period $t$ are the values of $b_t$ that maximize $U_t(b_t | y^t) = U_t(b_t | y^t)$. We now show that there exists a unique value $b^*_t$ of $b_t$ that maximizes $U_t(b_t, | y^t)$ regardless of $y^t$. First notice, by Lemma 2, that

$$U_t(b_t | y^t) \propto -\exp\left(-rM_t + \frac{1}{2}r^2V_t\right),$$

and so the values of $b_t$ that maximize $U_t(b_t | y^t)$ are independent of $y^t$. We are done if we show that $U_t(b_t | y^t)$ has a unique maximizer. Let $h_t(b_t) = a^*_t(b_t) - g(a^*_t(b_t)) + \sum_{s=t+1}^{T} \delta^{t-s}k(a^*_t(b_t))$.

From (25), we have that

$$\frac{\partial a^*_t(b_t)}{\partial b_t} = \frac{1}{g''(a^*_t(b_t)) - \sum_{\tau=t+1}^{T} \delta^{t-\tau}B^*_\tau k''(a^*_t(b_t))}.$$

Hence, using (25) one more time,

$$h_t'(b_t) = \frac{\partial a^*_t(b_t)}{\partial b_t} \left[ 1 - B_t - \sum_{\tau=t+1}^{T} \delta^{t-\tau}(1 - B^*_\tau)k'(a^*_t(b_t)) \right],$$

from which we obtain that

$$h_t''(b_t) = -\frac{\partial a^*_t(b_t)}{\partial b_t} \left[ \frac{g'''(a^*_t(b_t)) - \sum_{\tau=t+1}^{T} \delta^{t-\tau}B^*_\tau k'''(a^*_t(b_t))}{\left[ g''(a^*_t(b_t)) - \sum_{\tau=t+1}^{T} \delta^{t-\tau}B^*_\tau k''(a^*_t(b_t)) \right]^2} \right. \left. - \frac{\partial a^*_t(b_t)}{\partial b_t} \left[ 1 + \sum_{\tau=t+1}^{T} \delta^{t-\tau}(1 - B^*_\tau)k''(a^*_t(b_t)) \frac{\partial a^*_t(b_t)}{\partial b_t} \right] \right].$$

Thus, since $\sum_{\tau=t+1}^{T} \delta^{t-\tau}B^*_\tau \geq 0$ if $k$ is non–linear by (iv) in the induction hypothesis, the function $h_t(b_t)$ is strictly concave. This implies that $U_t(b_t | y^t)$ is strictly quasi–concave, which in turn implies the desired result.

To summarize, we have established if $(\sigma^*, \omega^*)$ is to be an equilibrium, then it must be uncontingent in period $t$. Moreover, if $b^*_t$ is the piece rate in period $t$, condition (25) implies that the worker’s effort $a^*_t$ is the unique solution to

$$g'(a) = B^*_t + \sum_{\tau=t+1}^{T} \delta^{t-\tau}B^*_\tau k'(a),$$

44
where
\[ B_t^* = b_t^* + \sum_{s=t+1}^{T} \delta^{s-t}(1 - b_s^*)(1 - \mu_t) \prod_{\tau=t+1}^{s-1} \mu_\tau. \]

Since there is a one-to-one correspondence between \( b_t \) and \( B_t \), we can describe the equilibrium piece rate in period \( t \) in terms of the value \( B_t^* \) of \( B_t \) that maximizes \( U_t(B_t|y') \).

By Lemma 2, the necessary and sufficient first-order condition for this maximization problem is given by
\[ \frac{\partial a_t^*}{\partial B_t} \left[ 1 - g'(a_t^*(B_t)) + \sum_{\tau=t+1}^{T} \delta^{\tau-t}k'(a_t^*(B_t)) \right] = \left( B_t + \sum_{\tau=t+1}^{T} \delta^{\tau-t}B_\tau^* \right) r\sigma_t^2 + B_tr\sigma_\varepsilon^2. \]

By (25), the above equation reduces to
\[ \frac{\partial a_t^*}{\partial B_t} \left[ 1 - B_t + \sum_{\tau=t+1}^{T} \delta^{\tau-t}(1 - B_\tau^*)k'(a_t^*(B_t)) \right] = \left( B_t + \sum_{\tau=t+1}^{T} \delta^{\tau-t}B_\tau^* \right) r\sigma_t^2 + B_tr\sigma_\varepsilon^2. \]

Since \( \partial a_t^*/\partial B_t = \partial a_t^*/\partial b_t \), equation (26) implies that the last equation can be rewritten as
\[ H_t(B_t) = 1 - B_t - r \left[ g''(a_t^*(B_t)) - k''(a_t^*(B_t)) \sum_{\tau=t+1}^{T} \delta^{\tau-t}B_\tau^* \right] \left[ \Sigma_t^2 B_t^* + \sigma_t^2 \sum_{\tau=t+1}^{T} \delta^{\tau-t}B_\tau^* \right] + k'(a_t^*(B_t)) \sum_{\tau=t+1}^{T} \delta^{\tau-t}(1 - B_\tau^*) = 0. \]

(27)

We are done if we show that \( \sum_{\tau=t}^{T} \delta^{\tau-t+1}B_\tau^* \geq 0 \) when \( k \) is non-linear. Suppose not, that is, suppose that
\[ \sum_{\tau=t}^{T} \delta^{\tau-t+1}B_\tau^* = \delta B_t^* + \delta \sum_{\tau=t+1}^{T} \delta^{\tau-t}B_\tau^* < 0, \]
which implies that
\[ \sum_{\tau=t+1}^{T} \delta^{\tau-t}B_\tau^* < -B_t^*. \]
(28)

Now observe, by (27), that
\[ B_t^* = \frac{1 - r \left[ g''(a_t^*) - k''(a_t^*) \sum_{\tau=t+1}^{T} \delta^{\tau-t}B_\tau^* \right] \sigma_t^2 \sum_{\tau=t+1}^{T} \delta^{\tau-t}B_\tau^* + k'(a_t^*) \sum_{\tau=t+1}^{T} \delta^{\tau-t}(1 - B_\tau^*)}{1 + r\Sigma_t^2 \left[ g''(a_t^*) - k''(a_t^*) \sum_{\tau=t+1}^{T} \delta^{\tau-t}B_\tau^* \right]} \]
and so (28) is equivalent to

$$\sum_{\tau=t+1}^{T} \delta^{\tau-t} B^*_\tau \left\{ 1 + r \sigma_\varepsilon^2 \left[ g''(a^*_t) - k''(a^*_t) \sum_{\tau=t+1}^{T} \delta^{\tau-t} B^*_\tau \right] - k'(a^*_t) \right\} < -1 - k'(a^*_t) \sum_{\tau=t+1}^{T} \delta^{\tau-t}.$$ 

Since $k'(a) \leq 1 + r \sigma_\varepsilon^2 g''(a)$ for all $a$, we then have that $\sum_{\tau=t+1}^{T} \delta^{\tau-t} B^*_\tau \geq 0$ implies that the term in brackets on the left-hand side of the above equation is positive. This, however, implies that $\sum_{\tau=t+1}^{T} \delta^{\tau-t} B^*_\tau < 0$, a contradiction.