The Neoclassical Theory of the Firm
Under Moral Hazard

Michael T. Rauh
Indiana University
mtrauh@indiana.edu

April 6, 2017

Abstract

We develop a neoclassical theory of the firm under moral hazard with endogenous employment, endogenous capital, and an external competitive labor market. The crucial assumptions are that effort becomes harder to measure as employment grows and the exogenous parameters are affiliated. The model explains why incentives decline but wages rise with firm size, the mixed evidence on the risk-reward tradeoff, and the positive correlation between wages and profits. In the long run there is a positive relationship between incentives and risk driven by endogenous capital. Finally, the model makes novel predictions about the relationship between incentives and aggregate labor market conditions.

JEL Classifications: D02, D21, D86, J31, M52.

Keywords: employment, incentives, moral hazard, size-wage differential, rent-sharing, risk-reward tradeoff, wages.
1 Introduction

In this paper, we develop a neoclassical theory of the firm under moral hazard with endogenous employment, endogenous capital, and an external perfectly competitive labor market. Our purpose is twofold. First, we show that the model is capable of explaining the following important stylized facts about incentives and wages:

(SF1) Incentives decline with the size of the firm as measured by either employment or capital. Bishop (1987); Rasmussen and Zenger (1990); Garen (1994); Schaefer (1998); Zenger and Marshall (2000); and Zenger and Lazzarini (2004).

(SF2) The classical risk-reward tradeoff in the principal-agent literature implies that incentives should decline with risk but the evidence as surveyed in Prendergast (2002a) and Devaro and Kurtulus (2010) is mixed. Indeed, a majority of studies find a positive relationship between incentives and risk.

(SF3) The size-wage differential, which states that wages increase with firm size. Barron, Black, and Loewenstein (1987); Brown and Medoff (1989); Abowd, Kramarz, and Margolis (1999); Troske (1999); and many others.

(SF4) Rent-sharing, as manifested by a positive correlation between wages and profits. Blanchflower, Oswald, and Sanfey (1996); Arai (2003); and several others.

The model is an extension of Holmström and Milgrom (1987), which assumes linear contracts, exponential utility, and a normally distributed additive shock. In the short run, the firm can freely vary employment but the capital stock is fixed. The endogenous variables are therefore incentives, wages, employment, and profit while the exogenous parameters are price, total factor productivity, capital, the inverse of capital costs, and the inverse of risk. We specify the latter two parameters in terms of their inverses so we can state that optimal employment is increasing in all the parameters. We make two crucial assumptions. The first is that effort becomes harder to measure as employment grows, an idea that dates back at least to Stigler (1962) and has been formally analyzed in Garen (1985); Ziv (1993); Auriol, Friebel, and Pechlivanos (1999, 2002); Liang, Rajan, and Ray (2008);
and Rauh (2014). The second main assumption, following Holmström and Milgrom (1994), is that the parameters are affiliated, so a firm which has a high value for one parameter will tend to have high values for all of them. Since optimal employment is increasing in all the parameters, such a firm will tend to be a large employer which offers weak incentives because effort is hard to measure in large firms (SF1) but high wages as a compensating differential for risk (SF3). Since workers are compensated on the basis of output, which under certain conditions is correlated with profit, wages will be positively correlated with profit (SF4). In our model, what appears to be rent-sharing is actually a consequence of performance-related pay.

A substantial theoretical literature already exists for (SF2)-(SF4) and we discuss each of them more fully below. What sets our model apart from the existing theoretical literatures on the size-wage differential and rent-sharing is that most current explanations are based on external market considerations, whereas in this paper we offer a complementary internal explanation based on incentive contracting within the firm. Furthermore, most of these explanations rely on differences across workers or the existence of market power in product and/or labor markets, whereas in this paper we show that a size-wage differential and apparent risk-sharing can arise even when workers are identical and all markets are perfectly competitive. An important advantage of our approach is that it provides a comprehensive explanation for all four stylized facts within a single theoretical framework.

A concise summary of these theoretical results is provided by the classic empirical paper by Dickens and Katz (1986). Their paper seeks to explain substantial differences in wages across industries for otherwise seemingly identical workers and to identify those industry characteristics which explain these wage differences. Summarizing their findings from many different specifications, they conclude that three explanatory variables stand out: education (corresponding to total factor productivity), the capital-labor ratio, and the profit variables. The main obstacle to obtaining robust results is that the explanatory variables are highly correlated. After performing a principal component analysis, the authors find that the first factor accounts for over a third of the variance of the industry variables. This is the empirical expression of our assumption that the exogenous variables are affiliated and therefore tend to move together. The list of variables which are correlated with this one dominant factor includes wages, variables related to total factor productivity (labor

---

3This premise seems widely accepted in the literature. The only concrete evidence we are aware of is Troske (1999), who finds that the number of supervisors per employee is uncorrelated with both establishment and firm size. This implies that the number of supervisors increases proportionately with employment.

4Of course, there are other stylized facts the model cannot explain, such as the fact that larger firms have lower quit rates (e.g., see Bertola and Garibaldi (2001)).
productivity, education, and job tenure), firm and establishment size in terms of employment, the capital-labor ratio, and the profit variables. This is a rather striking empirical portrait of our theoretical results.

The other major objective of the paper is to provide a simple theoretical framework which incorporates moral hazard but has the same broad scope of application as the textbook competitive model. In particular, situating the firm within a competitive labor market allows us to move beyond the above stylized facts to make new predictions about the relationship between incentives and aggregate labor market conditions:

(P1) A leftward shift in the supply of labor (e.g., a reduction in labor force participation) will lead to stronger incentives at all firms.

(P2) We show that the short-run optimal incentive does not depend on firm risk (SF2), but a leftward shift in the aggregate demand for labor due to a market-wide increase in risk leads to weaker incentives at all firms. In other words, there is no observable risk-reward tradeoff at the firm level but there is one at the market level.

More generally, we show that the optimal incentive is increasing in the workers’ outside option which is the highest alternative total payoff available in the labor market. Since the participation constraint binds, the total payoff of each worker is equal to the outside option which is determined by supply and demand in the labor market. Intuitively, an increase in the outside option (due to a rightward shift in demand or a leftward shift in supply) increases the cost of employment. The subsequent reduction in employment makes effort easier to measure so firms offer stronger incentives. Predictions (P1) and (P2) above are special cases of this general argument. In the principal-agent and teams literatures, the outside option is usually normalized to zero because all it does is determine the distribution of the surplus between the principal and agent(s). In contrast, in our model with endogenous employment the outside option is an important endogenous variable which firms take as given when choosing employment, incentives, and wages. Garen (1994) provides some preliminary evidence in support of (P2) but we are unaware of any evidence on (P1).

In the long run, firms are free to adjust their capital stocks and we can add capital to the previous list of endogenous variables. The fact that firms in our model make explicit capital decisions leads to further new predictions.

(P3) An increase in the cost of capital will lead to stronger incentives.

(P4) In the long run there is a positive relationship between incentives and risk.
Prediction (P4) may seem counterintuitive from the standpoint of conventional contract theory but is consistent with some evidence as mentioned in the context of (SF2). In our model, an increase in the cost of capital (P3) or an increase in risk (P4) reduces the demand for capital which in turn reduces the demand for labor because employment and capital are complements in the production function. Again, the reduction in employment makes effort easier to measure and leads to stronger incentives. In both cases, the initial impact on incentives operates through the endogenous capital stock so these are inherently long-run effects.

In the formal model we follow Holmström (1982) and assume incentive contracts based on group performance and show that contracts based instead on individual performance do not necessarily change the results. In fact, the direct nature of the intuitions suggests that our results should hold in any model where earnings depend on measured performance and measuring performance gets more difficult as employment increases. This includes not only explicit incentives tied to individual or group performance but also implicit incentives and perhaps other incentive mechanisms such as promotion tournaments.

The plan for the rest of the paper is as follows. In the next section we lay out the model primitives. In section 3, we consider the short run where the capital stock is fixed. We consider the long run in section 4. Section 5 concludes.

2 Model Primitives

We consider a labor market where the supply side consists of infinitely many identical workers. The demand side consists of \( n \) firms, where \( n \) is large enough to justify perfect competition in the labor market. These firms all hire labor from the same labor pool but may operate in different product markets, producing different outputs with different production technologies. All markets are perfectly competitive but we only explicitly model equilibrium in the labor market.

We now describe the production function. Let \( L_i \) be the number of workers employed by firm \( i \). Each worker \( j \) at firm \( i \) chooses an effort level \( e_{ij} \) (we discuss this choice below) and experiences an idiosyncratic random shock \( \epsilon_{ij} \). Total effort at firm \( i \) is given by

\[
E_i = \sum_{j=1}^{L_i} e_{ij}
\]

and the firm shock by

\[
\epsilon_i = \sum_{j=1}^{L_i} \epsilon_{ij}.
\]
For each firm $i$ the random variables $\{\epsilon_{ij}\}$ are i.i.d. normal across workers with mean zero and variance $\sigma_i^2$, so the firm shock $\epsilon_i$ is normal with mean zero and variance $\sigma_i^2 L_i$. Note that greater employment $L_i$ increases the volatility of the firm shock. Since workers are identical, the variance $\sigma_i^2$ is the same for all workers but may vary across firms depending on their production technology.

The stochastic production function of firm $i$ is given by

$$q_i = A_i f(K_i) E_i + \epsilon_i,$$  \hfill (3)

where $q_i$ is output, $A_i > 0$ is total factor productivity, $K_i$ the capital stock, and $f$ satisfies $f(0) = 0$, $f' > 0$, and $f'' < 0$. The deterministic part of (3) is essentially Cobb-Douglas in capital $K_i$ and total effort $E_i$. The total factor productivity parameter $A_i$ captures investments in human capital, process improvements such as employee involvement programs, and broader aspects of technological progress such as specialization and division of labor (none of which are modeled here). The function $f$ reflects the productivity of capital. The stochastic part $\epsilon_i$ represents the crucial assumption that each additional worker adds an additional productivity shock. For example, consider an assembly line where a separate workstation is created for each new worker. This adds one more stage in the production process and therefore one more source of uncertainty. Another interpretation, discussed below, is that each additional worker adds additional measurement error in group performance measurement.

We assume moral hazard in the sense that the firm observes output $q_i$ but not total $E_i$ or individual efforts. As the size of the firm in terms of employment $L_i$ grows, the variance of output $\sigma_i^2 L_i$ increases and output becomes a noisier signal of effort. In this sense, effort becomes harder to measure. We refer to $\sigma_i^2$ as the marginal variance because it represents the increase in the variance of output due to the marginal hire. The parameters $A_i$ and $\sigma_i^2$ characterize the production technology of firm $i$ and summarize technological differences across firms.

For most of the paper we follow Holmström (1982) and assume that individual performance measures are not available and that output (3) is the only contractible performance measure.\footnote{A performance measure is contractible if it is observable to the parties to the contract and verifiable to third parties such as the courts. It can therefore be included as part of an enforceable contract.} In particular, firm $i$ offers a linear contract in output

$$I_i = \alpha_i + \beta_i q_i = \alpha_i + \beta_i A_i f(K_i) E_i + \beta_i \epsilon_i,$$ \hfill (4)

where $I_i$ is the wage or income, $\alpha_i$ the fixed component of compensation, and $\beta_i$ the incentive
parameter. We discuss individual performance measures briefly below. Note that incentives tied to stochastic output expose workers to risk. In particular, the effect on income (4) of fluctuations in the firm shock $\epsilon_i$ is magnified by the incentive $\beta_i$ as is evident from the final term.

**Signal Interpretation**

In the above interpretation of the model, workers are compensated based on actual output (3) which is subject to productivity shocks $\epsilon_i$. We call this the *output* interpretation of the model. Alternatively, we could assume that output is deterministic

$$q_i = A_i f(K_i) E_i$$

but that firms must pay their workers before output is realized based on a noisy signal of output

$$y_i = q_i + \epsilon_i.$$  \hspace{1cm} (6)

In this interpretation of the model, $\epsilon_i$ represents noise, so each additional hire increases noise or measurement error. The corresponding contract is

$$I_i = \alpha_i + \beta_i y_i.$$  \hspace{1cm} (7)

We call this the *signal* interpretation of the model. These two interpretations of the model are obviously equivalent but their empirical implications differ. In the output version, a reduction in the marginal variance $\sigma_i^2$ can be interpreted in terms of quality control or process improvements, whereas in the signal version it can be interpreted as an increase in monitoring which improves the informativeness of the signal. For concreteness, we proceed with the output interpretation because it involves less notation. □

Let $M_i$ be the subset of the $n$ firms which operates in the same product market as firm $i$ and $p_{M_i}$ the perfectly competitive price determined by supply and demand in that market. Let $\rho_i$ be the cost of capital, which reflects the cost of borrowing for firm $i$. Since firms can differ in terms of

---

6As is well-known (see Bolton and Dewatripont (2005, Section 4.3)), linear contracts are generally suboptimal in the context of additive and normally distributed productivity shocks and a single agent with CARA utility. Given the structure of our model, in particular the additive nature of the production function, there is reason to believe that the classic rationale for linear contracts provided by Holmström and Milgrom (1987) for the single-agent context also applies here. Under limited liability, Bose, Pal, and Sappington (2011) show that the optimal linear contract achieves at least 90% of the expected profit of the optimal nonlinear contract. A linear contract will therefore be optimal when transaction costs are sufficiently increasing in contractual complexity.
price $p_{M_i}$, total factor productivity $A_i$, capital $K_i$, and marginal variance $\sigma_i^2$, their creditworthiness and hence their borrowing costs will also generally differ. The expected wage $w_i$ and wage $I_i$ are given by

$$w_i = \alpha_i + \beta_i A_i f(K_i) E_i$$

$$I_i = w_i + \beta_i \epsilon_i$$

and profit $\bar{\Pi}_i$ and expected profit $\Pi_i$ by

$$\bar{\Pi}_i = p_{M_i} q_i - I_i L_i - \rho_i K_i$$

$$\Pi_i = p_{M_i} A_i f(K_i) E_i - w_i L_i - \rho_i K_i.$$  

Workers have the CARA utility function

$$-\exp\left\{-r \left[ I_i - (1/2)\epsilon_{ij}^2 \right]\right\},$$

where $r$ is the CARA coefficient and $(1/2)\epsilon_{ij}^2$ the cost of effort. A higher value of $r$ indicates that workers are more risk averse. A worker’s payoff at firm $i$ is given by\footnote{For a derivation, see Bolton and Dewatripont (2005, p. 137). Since all workers at firm $i$ choose the same effort level $e_{ij}$, we write $U_i$ instead of $U_{ij}$.}

$$U_i = \alpha_i + \beta_i A_i f(K_i) E_i - (1/2)\epsilon_{ij}^2 - RP_i,$$

where

$$RP_i = (1/2)r \beta_i^2 \sigma_i^2 L_i$$

is the workers’ risk premium which represents the disutility of risk. This is essentially the variance of income $\beta_i^2 \sigma_i^2 L_i$ scaled by the workers’ degree of risk aversion $r$. The risk premium is increasing in the incentive $\beta_i$ because stronger incentives increase the variance of income. Let $\bar{u}_i$ be the best alternative payoff for workers at firm $i$ across the other $n - 1$ firms which participate in this labor market

$$\bar{u}_i = \max_{k \neq i} \bar{U}_k.$$ 

Perfect competition will ensure that all firms will offer the same total payoff $U_i$ so we usually write $\bar{u}$ without the subscript.
3 The Short Run

The short run is defined as that period of time when firms’ capital stocks are fixed. We first consider the workers’ optimal choice of effort. Each employed worker \( j \) at firm \( i \) chooses effort \( e_{ij} \) to maximize her payoff (13). The solution is

\[
e_i = \beta_i A_i f(K_i)
\]

(16)

for all workers \( j \), so we drop the unnecessary subscript \( j \). As usual, stronger incentives \( \beta_i \) inspire greater effort because they increase the reward for additional output. But incentives are not the only factor which determines effort. An increase in total factor productivity \( A_i \) or the capital stock \( K_i \) also inspires greater effort because they increase the effectiveness of effort in generating output, which is the basis for rewards. Furthermore, an increase in \( A_i \) or \( K_i \) increases the sensitivity of effort to incentives. Note that these effects of \( A_i \) and \( K_i \) are conditional on the existence of some form of incentive pay. Otherwise, an increase in the expected marginal product of effort would not induce greater effort because output is not rewarded.

The short-run problem of firm \( i \) is to choose the contract \( (\alpha_i, \beta_i) \) and employment level \( L_i \geq 0 \) to maximize expected profit (11) subject to two constraints: the incentive compatibility constraint (16) and the participation constraint \( U_i \geq \bar{\pi} \). The incentive compatibility constraint captures the moral hazard problem internal to the firm while the participation constraint reflects the external competitive pressure on internal incentive contracting. It is clear that the firm will choose \( \alpha_i \) to make the participation constraint bind. Substituting \( U_i = \bar{\pi} \) or

\[
w_i = (1/2)e_i^2 + (1/2)r\beta_i^2 \sigma_i^2 L_i + \bar{\pi}
\]

(17)

into (11),

\[
\Pi_i = p_M A_i f(K_i) E_i - \left[ (1/2)e_i^2 + (1/2)r\beta_i^2 \sigma_i^2 L_i + \bar{\pi} \right] L_i - \rho_i K_i.
\]

(18)

The participation constraint (17) implies that workers must be compensated for an increase in effort, the risk premium (14), or the outside option \( \bar{\pi} \). Substituting optimal effort (16),

\[
\Pi_i = p_M A_i^2 f(K_i)^2 \beta_i L_i - \left[ (1/2)A_i^2 \beta_i^2 f(K_i)^2 + (1/2)r\beta_i^2 \sigma_i^2 L_i + \bar{\pi} \right] L_i - \rho_i K_i.
\]

(19)

We now derive the optimal incentive \( \beta_i \) for a given employment level \( L_i \). From (19), the expected benefit of incentives is that they increase expected revenue by increasing effort and expected output
(the first term). The cost is that incentives necessitate an increase in the expected wage \( w_i \) in (17) because of the higher cost of effort and the increase in the risk premium (the second and third terms). The optimal \( \beta_i \) for a given employment level \( L_i \) balances these tradeoffs

\[
\beta_i = \frac{p_M A_i^2 f(K_i)^2}{A_i^2 f(K_i)^2 + r \sigma_i^2 L_i}.
\]  

(20)

In comparison, the optimal incentive in Milgrom and Roberts (1992, p. 221) in a single agent context is given by

\[
\beta = \frac{P'(e)}{1 + r \sigma^2},
\]

(21)

where \( r \) and \( \sigma^2 \) are the same as in this paper and \( P(e) \) is the expected benefit of effort to the principal. Since \( P(e) \) presumably captures such effects as the price \( p_M \), and productivity \( A_i \), our expression (20) has a similar structure except that it also depends on capital \( K_i \) and employment \( L_i \). But \( \beta_i \) in (20) is not the optimal incentive in our model because employment is endogenous. After substituting the optimal employment level into (20), it will lose the familiar structure in (21).

In our model, an increase in the price \( p_M \) makes effort more valuable to the firm, which responds with stronger incentives \( \beta_i \). An increase in the CARA coefficient \( r \) or the variance \( \sigma_i^2 L_i \) of output increases the cost of incentives in terms of the risk premium which leads to weaker incentives. This is a version of the classical risk-reward tradeoff already evident in (21). In particular, an increase in employment \( L_i \) leads to weaker incentives by making effort harder to measure. Incentives are increasing in total factor productivity \( A_i \) and the capital stock \( K_i \) due to two separate effects. First, an increase in \( A_i \) or \( K_i \) makes effort more valuable to the firm. Second, an increase in \( A_i \) or \( K_i \) increases the sensitivity of effort to incentives in the incentive compatibility constraint (16). These two effects explain why \( A_i \) and \( K_i \) enter (20) as squared terms. Note that the effect of firm size on incentives is currently ambiguous: an increase in firm size as measured by employment \( L_i \) leads to weaker incentives whereas an increase in firm size as measured by assets \( K_i \) leads to stronger incentives. Once the endogeneity of employment \( L_i \) is taken into account, we will find that the optimal incentive is unambiguously decreasing in firm size as measured by either capital or employment.

3.1 The Demand for Labor and Equilibrium in the Labor Market

As we have seen, each firm \( i \) takes the workers’ outside option \( \bar{\pi} \) as given and chooses the fixed component of pay \( \alpha_i \) to make the participation constraint bind \( U_i = \bar{\pi} \). All workers at firm \( i \) receive the same payoff \( \bar{\pi} \) and in equilibrium this must be the same across all firms. Note that individual
firms choose employment \( L_i \) and employment contracts \((\alpha_i, \beta_i)\) while perfect competition in the labor market determines the total payoff \( \overline{u} \) of the workers. Compensation is therefore a product of both internal (moral hazard) and external (market competition) considerations.

The first step in the construction of a competitive labor market equilibrium is the demand for labor by an individual firm \( i \). In the present context, the usual “price-taking” assumption is replaced by a “payoff-taking” assumption where each firm believes that it can hire as many workers as it wants as long as it offers the market payoff \( \overline{u} \). The demand for labor by firm \( i \) will therefore be a function

\[
L_i(\pi | p_{Mi}, A_i, K_i, r, \sigma_i^2),
\]

where changes in \( \overline{u} \) correspond to movements along the curve while changes in the parameters \((p_{Mi}, A_i, K_i, r, \sigma_i^2)\) shift the curve.

The various costs and benefits of employment \( L_i \) given that the firm offers optimal incentives are evident from (17), (19), and (20). First, an increase in employment increases expected output and revenue. Second, an increase in employment leads to weaker incentives and therefore less effort. This reduces the risk premium and the cost of effort in (17), which lowers the necessary expected payment \( w_i \). Nevertheless, the expected wage bill \( w_i L_i \) rises. The optimal employment level balances these tradeoffs. Substituting \( \beta_i \) from (20) into (19),

\[
\Pi_i = \frac{p_{Mi} A_i f(K_i)^4}{2 [A_i f(K_i)^2 + r \sigma_i^2 L_i]} L_i - \overline{u} L_i - \rho_i K_i.
\]

**Proposition 1** The demand for labor by firm \( i \) is given by

\[
L_i(\overline{u} | p_{Mi}, A_i, K_i, r, \sigma_i^2) = \begin{cases} 
\frac{A_i f(K_i)^2 [p_{Mi} A_i f(K_i) - \overline{u}]}{r \sigma_i^2 \sqrt{2\pi}} & \text{if } p_{Mi} A_i f(K_i) \geq \sqrt{2\overline{u}} \\
0 & \text{otherwise.}
\end{cases}
\]

At positive employment levels, employment is increasing in the price \( p_{Mi} \), total factor productivity \( A_i \), and capital \( K_i \) and decreasing in the CARA coefficient \( r \), the marginal variance \( \sigma_i^2 \), and the outside option \( \overline{u} \).

In the expression (17) for the expected wage \( w_i \), the outside option \( \overline{u} \) represents the external or market cost of employment. At positive employment levels, the demand for labor is therefore downward-sloping in \( \overline{u} \) (with \( \overline{u} \) on the vertical axis and \( L_i \) on the horizontal). The internal costs of employment are the cost of effort and the risk premium which must be compensated to satisfy

---

8A function \( g \) is increasing if \( x > y \) implies \( g(x) > g(y) \) and nondecreasing if \( x > y \) implies \( g(x) \geq g(y) \).
the participation constraint. An increase in the CARA \( r \) or the marginal variance \( \sigma_i^2 \) raises the internal cost of employment which shifts the demand for labor to the left. An increase in the price \( p_M \) or total factor productivity \( A_i \) shifts demand to the right because they increase the value of employment in terms of expected output. Firms with greater capital \( K_i \) have a greater demand for labor because capital and labor are complements in the production function (3).

The aggregate demand for workers is the horizontal sum

\[
L^d(\pi \mid \{p_{M_i}\}_{i=1}^n, \{A_i\}_{i=1}^n, \{K_i\}_{i=1}^n, r, \{\sigma_i^2\}_{i=1}^n) = \sum_{i=1}^n L_i(\pi \mid p_{M_i}, A_i, K_i, r, \sigma_i^2).
\]  

(25)

Note that the short-run aggregate demand for labor does not depend on capital costs \( \{\rho_i\} \) but the long-run aggregate demand for labor will. The final ingredient is the supply \( L^s(\pi) \) of workers which we assume is increasing in \( \pi \).\(^9\) A competitive labor market equilibrium is a \( \pi \geq 0 \) such that supply equals demand

\[
L^s(\pi) = L^d(\pi \mid \{p_{M_i}\}, \{A_i\}, \{K_i\}, r, \{\sigma_i^2\}).
\]  

(26)

In Figure 1 below, we depict the aggregate demand for labor in the special case where there are 10 identical firms who compete in the same product market with the same production function, with \( p = 10, A = 10, f(K) = 5, r = 10, \) and \( \sigma = 1 \). From the expression (24) for optimal employment, the aggregate demand curve crosses the \( \pi \) axis when \( \pi = (1/2)A^2p^2f(K)^2 \) and then asymptotes to the \( L \) axis. Given the basic shape of aggregate demand, a unique equilibrium will exist for any well-behaved supply curve. Figure 1 depicts the unique equilibrium when \( L^s(\pi) = 10\sqrt{\pi} \).\(^{10}\)

---

\(^9\)We are therefore implicitly assuming that workers have different valuations of leisure or different employment opportunities outside the specific labor market which is our focus. Workers are otherwise identical.

\(^{10}\)For this particular supply curve, the equilibrium has a unique closed-form solution but it is not sufficiently interesting to report it here.
When markets are perfectly competitive, the choices and parameters which are specific to one firm have little or no effect on equilibrium prices. In our model, this means that the choices \((\alpha_i, \beta_i, L_i)\) and parameters \((A_i, K_i, \sigma_i^2)\) that are specific to firm \(i\) have a negligible impact on \(p_{Mi}\) and \(\overline{\pi}\). For example, an increase in productivity \(A_i\) or employment \(L_i\) at a single firm \(i\) will have no effect on the equilibrium payoff \(\overline{\pi}\) of the workers or the competitive price \(p_{Mi}\).

### 3.2 The Optimal Incentive

Substituting the optimal employment level (24) into the previous expression (20) for the optimal incentive given employment,

**Proposition 2** Given positive employment, the optimal incentive is given by

\[
\beta_i = \frac{\sqrt{2\sqrt{\pi A_i f(K_i)}}}{A_i f(K_i)},
\]

which is increasing in the outside option \(\overline{\pi}\) and decreasing in total factor productivity \(A_i\) and the capital stock \(K_i\).

The optimal incentive \(\beta_i\) depends on \(A_i\) and \(K_i\), which are specific to firm \(i\), as well as the workers’ equilibrium payoff \(\overline{\pi}\) which is the same for all firms and determined by supply and demand in the labor market. Comparing the optimal incentive given employment (20) with the optimal incentive (27), we observe that taking into account the endogeneity of employment has completely changed the nature of incentives. Previously, an increase in \(p_{Mi}\), \(A_i\), or \(K_i\) resulted in stronger incentives because these parameters make effort more valuable to the firm and an increase in \(A_i\) or \(K_i\) increases the sensitivity of effort to incentives. These results were consistent with the standard model (21). After substituting the optimal employment level, the optimal incentive (27) no longer depends on \(p_{Mi}\) and is decreasing in \(A_i\) and \(K_i\) which seems counterintuitive. But the previous result (20) was predicated on a fixed level of employment. When employment is endogenous, an increase in \(p_{Mi}\), \(A_i\), or \(K_i\) has the further effect, in addition to those above, of increasing employment which makes effort harder to measure and increases the cost of incentives in terms of the risk premium. In the case of \(p_{Mi}\) these two effects exactly offset, while in the case of \(A_i\) or \(K_i\) the second (negative) effect operating through employment dominates.

In most of the literature, which usually assumes a single agent, the outside option \(\overline{\pi}\) has no effect on incentives and is typically normalized to zero. Its only effect is on the fixed component of pay \((\alpha_i\) in our model) which determines the division of the total surplus between the principal and agent. But when employment is endogenous, an increase in \(\overline{\pi}\) increases the external cost of
employment, reduces employment, and makes effort easier to measure. This reduces the cost of incentives in terms of the risk premium and leads to stronger incentives.

The evidence for this prediction is sketchy and indirect but nevertheless encouraging. Bishop (1987, p. 548) finds that wages are more responsive to productivity in larger labor markets. His explanation is that “workers have a greater range of choices” which is broadly consistent with ours. In their survey of R&D engineers, Zenger and Lazzarini (2004) asked respondents to evaluate the degree to which their current job was “the only option available” and found that this was negatively correlated with pay mix (the ratio of bonuses to fixed pay) as well as perceived incentive intensity. Although neither of these findings directly addresses our prediction that incentives are increasing in the workers’ outside option, they are at least consistent with it.

3.3 The Holmström and Milgrom (1994) Framework

Some of the stylized facts we want to explain involve relationships between endogenous variables; e.g., the stylized facts that wages are increasing in profit and firm size. To generate empirical predictions of this sort we use the framework in Holmström and Milgrom (1994). Let $x$ and $y$ be random variables and $Cov(x, y)$ denote their covariance. We say that a vector $x$ of random variables is associated if $Cov[g(x), h(x)] \geq 0$ for all real-valued nondecreasing functions $g$ and $h$. It follows that the random variables $x$ have pairwise nonnegative covariances because $g$ and $h$ can be taken to be the appropriate projection maps. Intuitively, the concept of association captures a strong form of correlation which includes independence as a special case. According to Theorem 2(iv) in Holmström and Milgrom (1994), if $x$ is a vector of associated random variables and $g$ is a vector of nondecreasing real-valued functions then $(g(x), x)$ is associated. The vector $(g(x), x)$ will therefore exhibit nonnegative pairwise covariances.

3.4 Incentives and Firm Size

In the present context, we make the following assumption.

**Assumption 1** Given any values for the CARA coefficient $r$ and the payoff $\pi$ of the workers, let $T^s$ be the set of vectors

$$P^s_i = (p_{M_i}, A_i, K_i, 1/\sigma_i^2)$$

where the optimal employment level (24) and expected operating profit

$$p_{M_i}A_i f(K_i)E_i - w_iL_i$$

(29)
are nonnegative. We assume the vector $P_i^s$ of random variables is associated on $T^s$.

The expression in (29) is expected revenue minus expected variable cost (the expected wage bill) and the requirement that it be nonnegative is the standard short-run shutdown condition which appears in elementary textbooks. Given any values for $r$ and $\varpi$, $T^s$ is defined as the set of all parameter configurations $P_i^s$ such that employment is nonnegative and the firm does not shut down. The assumption is then that the random variables $P_i^s$ are associated on the set $T^s$.

This assumption implies that the parameters $P_i^s$ are positively correlated on $T^s$. This is a crucial assumption so we discuss its plausibility. In our model, $p_{Mi}$ is the perfectly competitive price determined by supply and demand in product market $M_i$ where firm $i$ operates. Since the parameters $(A_i, K_i, \sigma_i^2)$ specific to firm $i$ have no effect on $p_{Mi}$, there cannot be any correlation stemming from an influence of $(A_i, K_i, \sigma_i^2)$ upon $p_{Mi}$.

We now consider $A_i$ and $\sigma_i^2$, which characterize the production technology of firm $i$. One possibility, which seems quite plausible, is that these parameters are independent.\(^\text{11}\) Another is that firms’ production technologies can be ranked vertically, so that firms with superior technologies have high total factor productivity $A_i$ and low marginal variance $\sigma_i^2$. In this case, $A_i$ and $1/\sigma_i^2$ will be positively correlated. What the assumption rules out is negative correlation between $A_i$ and $1/\sigma_i^2$, which would imply that high-productivity firms tend to be high-volatility firms.

Finally, we consider the capital stock $K_i$ which is fixed in the short run. In the next section we show that the optimal long-run capital stock $K_i$ is nondecreasing in the vector $P_i^l$ of long-run parameters which includes $(p_{Mi}, A_i, 1/\sigma_i^2)$. If the parameters $P_i^l$ exhibit persistence across time then $K_{it}$, which was chosen optimally in the past based on $P_i^{l,t-1}$, should be positively correlated with $P_i^{l,t}$, which includes $(p_{Mi,t}, A_{i,t}, 1/\sigma_{i,t}^2)$.\(^\text{12}\) In that case, $P_i^s$ will be associated, where the strength of the positive correlations depends on the age of the capital stock and the extent of the persistence of the remaining parameters across time.

To apply Theorem 2(iv) in Holmström and Milgrom (1994), all the endogenous variables need to be nondecreasing in the parameters. Since the optimal incentive (27) is decreasing in total factor

\(^{11}\)Holmström and Milgrom (1994) assume independence in a similar context.

\(^{12}\)Formally, assume that $(P_i^{l,t-1}, P_i^{l,t})$ is associated which is a strong form of persistence. In the next section we show that $K_i$ is nondecreasing in the long-run parameters, so $(K_t(P_i^{l,t-1}), P_i^{l,t-1}, P_i^{l,t})$ is associated. Since any subvector of an associated vector of random variables is associated, $P_i^{l,t}$ is associated.
productivity $A_i$ and the capital stock $K_i$, we define the reciprocal as follows\textsuperscript{13}

$$
\tilde{\beta}_i(P^*_i) = \frac{1}{\beta_i(P^*_i)} = \frac{A_i f(K_i)}{\sqrt{2\sigma^2}}.
$$

We consider $\tilde{\beta}_i$ and the other endogenous variables to be functions solely of the parameters $P^*_i$ and not $(r, \bar{w})$ because we consider variations in the exogenous and endogenous variables with $(r, \bar{w})$ held fixed. Since optimal employment (24) and the reciprocal $\tilde{\beta}_i$ of the optimal incentive are nondecreasing in $P^*_i$ on the region $T^*$,

**Proposition 3** Given any values for $r$ and $\bar{w}$, the vector $(\tilde{\beta}_i(P^*_i), L_i(P^*_i), P^*_i)$ is associated on $T^*$.

This result applies purely at the individual firm level and does not involve any equilibrium considerations. Given any value $r$ for the CARA coefficient and any value $\bar{w}$ for the workers’ payoff (equilibrium or not), the exogenous variables $P^*_i$ and endogenous variables $\tilde{\beta}_i$ and $L_i$ should exhibit nonnegative pairwise covariances. In particular, the incentive $\beta_i$ should have nonpositive covariances with both capital $K_i$ and employment $L_i$. Since this result is set in the short run when capital stocks are fixed, these predictions are valid for cross-sectional data or panel data with a relatively short time horizon. Intuitively, firms that have a high value for one parameter will tend to have high values for all the parameters. Since employment is increasing in the parameters, such firms will tend to be large employers with large capital stocks which offer weak incentives because effort is hard to measure in large firms.

As far as we know, the only other paper which explicitly models the relationship between employment and incentives is Rauh (2014). In that paper, we consider the effects of specialization and division of labor in the context of Holmström and Milgrom (1987) and show that the relationship between employment and incentives can be positive or negative (Corollary 1). We also show that the expected wage is increasing in employment under certain conditions (Theorem 2).

All the available evidence supports the prediction that incentives are weaker in large employers. Rasmusen and Zenger (1990) find a positive relationship between wages and job tenure and that this relationship is stronger for large firms. Moreover, regressions of weekly earnings on tenure, outside experience, and education have a larger residual variance for small firms. These findings are consistent with the hypothesis in Garen (1985) and the present paper that performance is harder to

\textsuperscript{13}For notational simplicity and to conform with the formal statement of Theorem 2(iv) in Holmström and Milgrom (1994), we consider $\tilde{\beta}_i$ to be a function of $P^*_i$ despite the fact that it does not depend on $p_M$, or $\sigma^2$. We follow this practice throughout the paper. Note that $\tilde{\beta}_i$ is increasing in the parameters $A_i$ and $K_i$, which actually appear as arguments and nondecreasing in the rest.
measure in large firms, which therefore compensate on the basis of easily observed characteristics such as education and seniority, while small firms directly reward performance. Bishop (1987) provides direct evidence that wages increase with productivity for small but not for large firms. Zenger and Lazzarini (2004) find that the pay mix and subjectively perceived incentive intensity decline with firm size, albeit in a small survey of R&D engineers. Zenger and Marshall (2000) focus exclusively on group-based rewards (as in this section) and find that incentives decline with firm size for profit-sharing plans which operate at the firm level and that incentives decline with unit size for gain-sharing plans which operate at the unit level (i.e., division, facility, department, work group, and small teams).

The above evidence concerns the relationship between incentives and employment. Garen (1994) and Schaefer (1998) find evidence consistent with our other prediction that incentives decline with firm size as measured by capital. To explain their empirical findings, both authors develop agency models where the size of the firm (the book value of assets or market value) makes effort harder to measure in the same way that greater employment makes effort harder to measure in our model. In our model, the effect of capital on incentives operates through a different channel; i.e., through its complementarity with employment.

### 3.5 The Risk-Reward Tradeoff

The risk-reward tradeoff present in the optimal incentive with exogenous employment (20) is no longer evident in the optimal incentive with endogenous employment (27). Specifically, $\beta_i$ in (27) does not depend on the marginal variance $\sigma_i^2$. When employment is exogenous, an increase in $\sigma_i^2$ makes effort harder to measure and leads to weaker incentives as is evident from (20). But when employment is endogenous, an increase in $\sigma_i^2$ reduces employment which makes effort easier to measure. In our model, these two effects cancel. This result is a special case of Proposition 3 in Liang, Rajan, and Ray (2008), where the variance of the productivity shock is assumed to take the form $CL_i^\gamma \sigma^2$, where $C$ is a positive constant, $L_i$ is employment, $\gamma$ is a positive constant, and $\sigma^2$ is the base variance. In our model, this structure arises because the shock $\epsilon_i$ is the sum of the individual shocks $\epsilon_{ij}$ added by each worker.

Taking the log on both sides of (27),

$$\ln \beta_i = \ln \sqrt{2\pi} - \ln A_i - \ln f(K_i), \quad (31)$$

where $A_i$ and $K_i$ are specific to firm $i$ and in equilibrium $\bar{\pi}$ is the same across firms. Given
perfect data, the expression (31) should hold exactly and the marginal variance $\sigma_i^2$ should not have any explanatory power for incentives. In practice, we may have observable measures of human capital such as education, experience, and job tenure, but total factor productivity $A_i$ also captures unobservable human capital as well as process improvements, specialization and division of labor, and other technological factors related to productivity which may not be observed. Likewise, we may have a measure of the capital stock $K_i$ but not the productivity of capital $f(K_i)$ which appears in (31). In that case, a proxy for $\sigma_i^2$ may appear to have explanatory power in a regression with the incentive $\beta_i$ as the dependent variable when $\sigma_i^2$ is correlated with $A_i$ and $f(K_i)$. In fact, we may find a positive relationship between incentives and risk as predicted in Proposition 3. On the other hand, if $1/\sigma_i^2$ is independent of the other parameters, the correlation is weak, or capital stocks are sufficiently old, we may not observe any relationship at the firm level although we may observe one at the market level as shown below.

The absence of a risk-reward tradeoff at the firm level is consistent with the mixed nature of the evidence as surveyed in Prendergast (2002a). Indeed, the most consistent finding in the empirical literatures on franchising and sharecropping is that the risk-reward relationship is a positive one. After his survey of the empirical literature, Prendergast (2002a) develops a single-agent moral hazard model to explain such a positive relationship. In his model, when the principal is uncertain about which task the agent should perform, she optimally delegates that decision to the agent and offers incentives tied to output. When there is less uncertainty, the principal simply commands the agent to perform a particular task and then monitors effort. This can lead to a positive relationship between incentives and risk. Devaro and Kurtulus (2010) provide evidence in support of these predictions as well as a more recent survey of the empirical literature.

Other explanations include Prendergast (2002b), which predicts a positive relationship between incentives and risk when performance measurement is subjective and susceptible to favoritism and strategic monitoring. Guo and Ou-Yang (2006) show that the risk-reward tradeoff may not obtain when the agent controls both the mean and the variance of the performance measure. Later we will see that a positive relationship between incentives and risk emerges in our model in the long run when capital stocks are endogenous.

### 3.6 Wages and Expected Wages

Substituting optimal effort (16), the optimal incentive (27), and the optimal employment level (24) into the previous expression for the expected wage (17),
Proposition 4 The expected wage is given by

$$w_i = p_M A_i f(K_i) \sqrt{\bar{u}} \frac{\sqrt{2}}{\sqrt{2}} + \pi,$$

which is increasing in the price $p_M$, total factor productivity $A_i$, the capital stock $K_i$, and the outside option $\bar{u}$.

An increase in the price, total factor productivity, or capital leads to greater employment which makes effort harder to measure. This necessitates an increase in the expected wage as a compensating differential for the increase in the risk premium. The expected wage is also increasing in the outside option to satisfy the participation constraint. The model therefore predicts differences in expected wages across firms due to differences in the parameters $(p_M, A_i, K_i)$, which correspond to differences in product market conditions and factors which determine the productivity of labor. Note that these differences in expected wages will not induce search by workers because the payoff $\bar{u}$ is the same across firms.

3.7 The Size-Wage Differential

One of the most durable findings in the empirical labor literature is that wages are positively correlated with firm size as measured by employment. Once again we can apply the relevant theorem in Holmström and Milgrom (1994) since $w_i$ is nondecreasing in $P^s_i$ by Proposition 4 and the parameters $P^s_i$ are associated by Assumption 1.

Proposition 5 Given $r$ and $\pi$, the vector $(\tilde{\beta}_i(P^s_i), L_i(P^s_i), w_i(P^s_i), P^s_i)$ is associated on $T^s$.

We can therefore add the expected wage $w_i$ to the previous list of exogenous and endogenous variables which should exhibit nonnegative pairwise covariances. Furthermore,

Lemma 1 If $x_i$ is any random variable which is independent from the productivity shock $\epsilon_i$ then $\text{Cov}(I_i, x_i) = \text{Cov}(w_i, x_i)$.

Proof. Since $x_i$ and $\epsilon_i$ are independent,

$$\text{Cov}(I_i, x_i) = E(I_i x_i) - E(I_i)E(x_i) = E[(w_i + \beta_i \epsilon_i) x_i] - E(w_i)E(x_i)$$

$$= E(w_i x_i + \beta_i \epsilon_i x_i) - E(w_i)E(x_i) = E(w_i x_i) - E(w_i)E(x_i)$$

$$= \text{Cov}(w_i, x_i),$$
which completes the proof. ■

All the exogenous and endogenous variables are independent from the productivity shock, so we can also add the wage $I_i$ to the list of variables which have nonnegative pairwise covariances. Since the expected wage $w_i$ is increasing rather than merely nondecreasing in its arguments, we would expect positive and significant coefficients on capital $K_i$ and employment $L_i$ in cross-sectional regressions with the wage $I_i$ as the dependent variable. Intuitively, firms that are high in one parameter will tend to be high in all of them and will therefore tend to be large employers. Since effort is hard to measure in large firms, they will tend to offer weak incentives but high expected wages as a compensating differential for risk.

3.7.1 Empirical Evidence

In an influential paper, Brown and Medoff (1989) find a substantial size-wage differential which cannot be explained by differences in working conditions (e.g., less autonomy in larger firms), the threat of unionization, labor or product market power, or monitoring costs. They show that differences in worker quality explain about one-half of that differential while the other explanations account for little. In this paper we have assumed identical workers and perfect competition in the product and labor markets. Our model can therefore explain the size-wage differential without reference to worker quality or market power.

Similarly, Troske (1999) finds that wages are increasing in the capital-labor ratio, the skill of the workforce (e.g., the percentage of workers with at least a college degree), product market concentration, and the skill of managers. The prediction that wages should be increasing in the skill of the workforce comes from matching models such as Kremer (1993), which predict that workers will be matched with other workers of like skill and paid accordingly. As predicted by efficiency wage models, Troske finds that wages are negatively correlated with monitoring (the number of supervisors divided by employment). But only the capital-labor ratio and workforce skill explain the size-wage differential and both together account for only about half. The other factors have a direct impact on wages but are uncorrelated with employment; i.e., they have no indirect effect on wages operating through employment. It is precisely these indirect effects which have the potential to explain the size-wage differential.

In our model, the parameters that increase the expected wage (32) and do so operating through employment (24) are the price $p_{M_i}$, total factor productivity $A_i$, and the capital stock $K_i$. Now consider a wage equation which includes employment as an explanatory variable. Since optimal employment (24) includes $A_i$ and $f(K_i)$, adding capital $K_i$ and variables that capture measurable
human capital should reduce the explanatory power of employment and therefore explain a portion of the size-wage differential. To explain the remainder we would need to include all important aspects of total factor productivity and the productivity of capital. This discussion is broadly consistent with Troske’s findings on the capital-labor ratio and various measures of human capital and workforce skill. Up to this point, the empirical literature has focused mainly on product market concentration and other measures of market power with mixed success. In our model firms are perfectly competitive and it is the price $p_{M_i}$ and other variables related to profitability which contribute to the size-wage differential. We will address this issue further in the context of the literature on rent-sharing.

### 3.7.2 Theoretical Explanations

A variety of theoretical explanations have been put forward to explain the size-wage differential. In this discussion we focus on more recent contributions and contributions which have some overlap with this paper. Hamermesh (1980) attributes the size-wage differential to the fact that larger firms tend to be more capital-intensive, which is complementary with worker quality. It follows that larger firms will employ higher-quality workers and therefore pay higher wages. As discussed earlier, the capital-labor ratio does indeed explain part of the size-wage differential but differences in worker quality only explain about one-half of it. Similarly, our model predicts that wages should be correlated with investments in capital and that some of this effect operates through employment. The mechanism, however, is different. In our model, workers are identical. An increase in capital induces an increase in employment because they are complements in production. This increase in employment leads to an increase in the expected wage because of the increase in the risk premium which necessitates a compensating differential.

Zábojník and Bernhardt (2001) develop a tournament model which captures some of the ideas in Hamermesh (1980). In their model, firms with superior production technologies or higher prices hire more workers. Greater employment translates into more competitive promotion tournaments and therefore greater investments in human capital. Tournament winners must be paid a premium to prevent poaching by other firms and since the expected productivity of the winner is increasing in the size of the tournament, so is the wage that must be paid to winners. In summary, more profitable firms hire more workers, who make greater investments in human capital, and are paid higher wages.

Another potential explanation is monopsony power. In a perfectly competitive labor market, the supply of labor to an individual firm is perfectly elastic and there is no relationship between
wages and firm size. But if employers have market power in the labor market they will face an upward-sloping supply curve and larger firms will have to pay higher wages. As Green, Machin, and Manning (1996) note, the supply of labor to an individual firm can be less than perfectly elastic for a variety of reasons. One reason could be search frictions, as in Burdett and Mortensen (1998).\textsuperscript{14} Green \textit{et al.} use that model to show that the size-wage differential should be increasing in the equilibrium level of profit and find evidence in support of that prediction. Note that our model also features a positive relationship between wages and employment but this is not an upward-sloping supply curve in the traditional monopsony sense. Instead, in our model wages increase with employment operating through the risk premium.

The class of models most similar to ours is the class of efficiency wage models, especially the version in Mehta (1998).\textsuperscript{15} From our perspective, the main insight of this literature is that if the probability of detecting shirking declines with firm size (or the cost of monitoring increases) then larger firms will pay higher wages (and monitor less). These models can explain the finding in Troske (1999) (as well as other papers) that monitoring and wages are negatively correlated. On the other hand, Troske also finds that monitoring does not explain the size-wage differential because it is uncorrelated with employment. Finally, the same prediction seems to be at odds with the evidence that incentives decline with firm size.

Rent-sharing models explain the size-wage differential as follows. Firms with a technological or product market advantage will tend to be larger and more profitable. If workers are somehow able to capture some of this surplus then larger firms will pay higher wages. The class of matching models in Bertola and Garibaldi (2001) can be thought of in this way. In their model, the endogenous relationship between wages and employment at the firm level is a negative one. What produces the size-wage differential in aggregate data is the interaction between wages, employment, and firm-specific random shocks. In particular, a positive shock to the workers’ marginal revenue product will increase both employment and profitability. If the total surplus is split between firms and workers in a fixed way, then greater expected surplus will translate into higher wages. The prediction that wages and profits should be positively correlated finds support in Blanchflower, Oswald, and Sanfey

\textsuperscript{14}Shi (2002) explains the size-wage differential using a combination of labor and product market power in a model of directed search in both markets. Each worker produces one unit of output, so a firm’s employment level fixes its capacity. In the product market, consumers can only visit one firm due to high search costs. Firms post prices and then allocate output randomly among visiting consumers. These consumers are therefore willing to trade off a higher price in return for a higher probability of obtaining the good. Since that probability is increasing in the employment level of the firm (i.e., capacity), large firms can charge high prices and are therefore willing to offer high wages. A size-wage differential obtains when demand is moderate and persists because of search frictions in the labor market.

\textsuperscript{15}Our model belongs to the class of principal-agent models characterized by a tradeoff between efficiency and insurance. In contrast, the basic efficiency wage model is isomorphic to the basic principal-agent model characterized by a tradeoff between efficiency and limited liability rent extraction. See Laffont and Martimort (2002, p. 174).
(1996) and Arai (2003) as well as several other studies. Most of these papers explain the size-wage differential on the basis of external market forces. The contribution of our paper is to complement the existing theoretical literature by showing that their results are re-enforced when we look inside the firm to consider the effects on wages which arise from incentive contracting. One advantage of our approach is that our model can explain a wide variety of durable empirical phenomena in addition to the size-wage differential.

3.8 Rent-Sharing

We now show that under certain conditions our model implies a nonnegative correlation between wages and profits, consistent with the empirical rent-sharing literature. Consider the expression (23) for the expected profit $\Pi_i(P_s)$ of firm $i$, where $L_i$ is the optimal employment level. We can apply the envelope theorem to conclude that $\Pi_i(P_s)$ is increasing in all the parameters $P_s$ except possibly the capital stock $K_i$. In the short run, $K_i$ is fixed and not necessarily optimal. If $K_i$ is less than the optimal capital stock then an increase in $K_i$ will increase $\Pi_i$ but if $K_i$ exceeds the optimum then an increase in $K_i$ will decrease $\Pi_i$. Since $\Pi_i$ is not increasing in all the parameters, we cannot apply the framework in Holmström and Milgrom (1994). Intuitively, the problem is that firms with excessive capital stocks will offer relatively high wages (32) but will have low expected profit. If, however, we can assume that $T_s$ in Assumption 1 is a lattice and that the vector $P_s$ of parameters is affiliated then we have the following result.\footnote{Assume $X$ is a lattice. A vector $x = (x_1, x_2)$ of random variables is affiliated on $X$ if it is associated on every sublattice of $X$. In particular, if $x$ is affiliated then the subvector $x_1$ is associated for any fixed $x_2 = \bar{x}_2$ because $\{x \mid x_2 = \bar{x}_2\}$ is a sublattice. If $g$ is a vector of real-valued functions which are nondecreasing in $x_1$ (but not necessarily $x_2$) then $(g(x), x_1)$ is associated for any fixed $x_2 = \bar{x}_2$. See Holmström and Milgrom (1994, p. 980-1) for further discussion.}

Proposition 6 If the vector $P_s$ of parameters is affiliated on $T_s$ then given $r$ and $\pi$

$$\left(\tilde{\beta}_i(P_s), L_i(P_s), w_i(P_s), \Pi_i(P_s), \bar{\Pi}_s\right)$$

is associated conditional on $K_i = K_i$, where $\bar{P}_s = (p_{M_i}, A_i, 1/\sigma_i^2)$.

For any fixed value of the capital stock $K_i$, the endogenous variables $(\tilde{\beta}_i, L_i, w_i, \Pi_i)$ and the remaining exogenous variables $\bar{P}_s$ will exhibit nonnegative pairwise covariances. This includes, in particular, the expected wage $w_i$ and expected profit $\Pi_i$. Since $w_i$ is increasing in its arguments, we would expect a positive and significant coefficient on $\Pi_i$ in a regression with $w_i$ as the dependent variable and which includes controls for capital $K_i$. Adding explanatory variables which capture
aspects of prices \( p_{Mi} \) and total factor productivity \( A_i \) may weaken the empirical relationship between \( w_i \) and \( \Pi_i \) but will not eliminate it as long as the regressors fail to adequately capture these parameters. Note that the main empirical results in Blanchflower et al. (1996, Table 2) are based on industry averages of wages and profits. Their main finding that average wages are positively related to average profit is therefore consistent with the above result. Also note that Blanchflower et al. obtain this result without controlling for capital or the capital-labor ratio. This is to be expected when the number of firms with excessive capital stocks is sufficiently small so that controlling for capital becomes unnecessary.

We now turn to the relationship between wages \( I_i \) and profit \( \tilde{\Pi}_i \) as opposed to the relationship between their expectations as discussed above. According to Lemma 1, the covariance between the wage \( I_i \) and any random variable \( x_i \) which is independent of the productivity shock \( \epsilon_i \) is the same as the covariance between the expected wage \( w_i \) and \( x_i \). The problem is that when we consider the relationship between wages \( I_i \) and profit \( \tilde{\Pi}_i \), the latter does depend on \( \epsilon_i \). Instead, we have the following result.

**Lemma 2**

\[
\text{Cov}(I_i, \tilde{\Pi}_i) = \text{Cov}(w_i, \Pi_i) + \sigma_i^2 E[\beta_i(p_{Mi} - \beta_iL_i)].
\] (37)

**Proof.** From (10),

\[
\tilde{\Pi}_i = \Pi_i + \epsilon_i(p_{Mi} - \beta_iL_i).
\] (38)

Since \( \epsilon_i \) is independent from all the other parameters,

\[
\text{Cov}(I_i, \tilde{\Pi}_i) = E(I_i\tilde{\Pi}_i) - E(I_i)E(\tilde{\Pi}_i)
\] (39)

\[
= E[(w_i + \beta_i\epsilon_i)(\Pi_i + \epsilon_i(p_{Mi} - \beta_iL_i))] - E(w_i)E(\Pi_i)
\] (40)

\[
= E[w_i\Pi_i + \beta_i\epsilon_i\Pi_i + w_i\epsilon_i(p_{Mi} - \beta_iL_i) + \epsilon_i^2\beta_i(p_{Mi} - \beta_iL_i)] - E(w_i)E(\Pi_i)
\] (41)

\[
= E(w_i\Pi_i) - E(w_i)E(\Pi_i) + \sigma_i^2 E[\beta_i(p_{Mi} - \beta_iL_i)]
\] (42)

\[
= \text{Cov}(w_i, \Pi_i) + \sigma_i^2 E[\beta_i(p_{Mi} - \beta_iL_i)],
\] (43)

which completes the proof. ■

Consider a positive shock which increases output by \( \epsilon_i \). From the worker’s perspective, this increases the wage \( I_i \) by \( \beta_i\epsilon_i \). From the firm’s perspective, it increases revenue by \( p_{Mi}\epsilon_i \) but also increases costs by \( L_i\beta_i\epsilon_i \). For a sufficiently large firm, the overall impact on profit \( \tilde{\Pi}_i \) will be negative. A positive productivity shock therefore increases wages \( I_i \) but may reduce profit \( \tilde{\Pi}_i \),
introducing a negative correlation in the relationship between the two, which is therefore weaker than the unambiguously positive relationship between expected wages $w_i$ and expected profit $\Pi_i$. Nevertheless, Arai (2003, Tables 3 and 4) finds a positive relationship between $I_i$ and $\Pi_i$ controlling for the capital-labor ratio. In our model, this occurs when the incentive $\beta_i$ or marginal variance $\sigma_i^2$ is sufficiently small or the covariance $\text{Cov}(w_i, \Pi_i)$ is sufficiently large.

In traditional rent-sharing models, such findings are explained in terms of bargaining, fairness, and similar considerations. In our model, rents are shared via incentive contracts which reward workers based on group performance which is positively correlated with profit. Our model therefore suggests that a portion of the “rents” obtained by workers in profitable industries can be explained by performance-related pay, which need not be explicit.

### 3.9 Common Shocks

We can also use the above approach to incorporate common shocks. For example, consider the case where total factor productivity $A_i = A_{Mi} + \gamma_i$ can be decomposed into a common shock $A_{Mi}$ which affects all firms in product market $M_i$ and an idiosyncratic shock $\gamma_i$ specific to firm $i$. Let $p_{Mi}$ be the perfectly competitive price in product market $M_i$ determined by supply and demand. This will be a function $p_{Mi}(A_{Mi}, \zeta_i)$ of the demand parameters $\zeta_i$ as well as the common productivity shock $A_{Mi}$ but it will not depend on parameters specific to firm $i$ such as $\gamma_i$. Unfortunately, $p_{Mi}$ is decreasing in $A_{Mi}$ because it shifts the aggregate supply curve in the product market to the right. This upsets the monotonicity we need. If, however, we can assume that $p_{Mi}$ is increasing in $\zeta_i$ and the new vector of parameters

$$\hat{P}_i^s = (\zeta_i, A_{Mi}, \gamma_i, K_i, 1/\sigma_i^2)$$

is affiliated then the endogenous variables

$$\left(p_{Mi}(\hat{P}_i^s), \beta_i(\hat{P}_i^s), L_i(\hat{P}_i^s), w_i(\hat{P}_i^s), \Pi_i(\hat{P}_i^s)\right)$$

and the remaining exogenous variables $(\zeta_i, \gamma_i, 1/\sigma_i^2)$ will be associated for each fixed value for the capital stock $K_i$ and the common shock $A_{Mi}$. In this case, the relevant regressions would have not control not only for capital stocks but also include industry and/or sector dummy variables to control for common shocks.
3.10 Labor Market Effects

Up to this point, our analysis has not involved any equilibrium considerations. We now consider the effects of shifts in aggregate demand and supply in the labor market. Let \( L = \sum_{i=1}^{n} L_i \) denote total employment.

**Corollary 1**

(i) An increase in price \( p_{M_i} \) for all firms will result in an increase in incentives \( \beta_i \), expected wages \( w_i \), and total employment \( L \).

(ii) An increase in the marginal variance \( \sigma_i^2 \) for all firms or an increase in the CARA coefficient \( r \) will reduce incentives, expected wages, and total employment.

(iii) If the supply of labor shifts to the left, incentives and expected wages will rise while total employment will fall.

All of these effects can be read from the expressions for optimal employment (24), incentives (27), and the expected wage (32). For example, consider an increase in price for all firms. While an increase in the price \( p_{M_i} \) specific to firms operating in product market \( M_i \) may or may not affect the aggregate demand for labor, depending on the number of such firms, an increase in all prices surely will. The *direct* effect of the general price increase is to increase employment (24) and expected wages (32) with no effect on incentives (27). The *indirect* effects all follow from the increase in the aggregate demand for labor, which increases the expected payoff \( \pi \) of the workers as well as total employment \( L \). The increase in \( \pi \) leads to an increase in incentives and a further increase in expected wages. Total employment \( L \) rises but employment \( L_i \) at firm \( i \) can rise or fall depending on the specific parameters.\(^{17}\)

As we have seen, the optimal incentive (27) does not depend directly on the marginal variance \( \sigma_i^2 \) so there is no direct relationship between incentives and risk at the *firm* level. There is, however, a risk-reward tradeoff at the *market* level in the sense that an increase in \( \sigma_i^2 \) for all firms leads to weaker incentives. The only direct effect of an increase in \( \sigma_i^2 \) for all \( i \) is to reduce employment \( L_i \) for all \( i \). The indirect effects stem from the shift in the aggregate demand for labor to the left, which reduces \( \pi \) and total employment \( L \). The decrease in \( \pi \) leads to weaker incentives and lower expected wages at all firms.\(^{18}\) Once again, total employment \( L \) falls but the effect on employment \( L_i \) at firm

---

\(^{17}\)The increase in \( \pi \) will have feedback effects in all the relevant product markets, shifting market supply curves to the left, resulting in further increases in \( p_{M_i} \) for all firms. Assuming stability, the end result will be a higher expected payoff \( \pi \) for the workers and higher prices \( p_{M_i} \) for all firms, so the above comparative statics remain valid allowing for these general equilibrium effects.

\(^{18}\)The rightward shift of the market supply curves in the product markets reinforces these comparative statics.
i is ambiguous, depending on the parameters. The only evidence for this prediction seems to be Garen (1994), who considers the ratio of firm R&D expenditures to book value of assets, averaged across the relevant industry. As predicted, Garen finds a negative relationship between incentives and industry risk.\(^\text{19}\)

The final result (iii) makes some novel connections between contract theory and broader social trends. For example, consider a reduction in the labor force participation rate. This has no direct effects, but the leftward shift of the labor supply curve leads to an increase in \(\overline{u}\) and a reduction in total employment \(L\). Not surprisingly, this leads to higher expected wages. More interestingly, it also leads to stronger incentives as firms substitute effort from the remaining workers to compensate for lower employment. We are unaware of any evidence on this point.

### 3.11 Individual Incentives

For simplicity, we have assumed that incentives are tied to group performance \(q_i\) but there is little reason to believe that incentives for individual performance would qualitatively change the results. Consider the signal interpretation of the model, where (5) is deterministic output. The individual output of worker \(j\) at firm \(i\) is given by

\[
q_{ij} = A_if(K_i)e_{ij},
\]

where \(q_i = \sum_j q_{ij}\). Assume the only contractible performance measure is a signal \(y_{ij} = q_{ij} + \epsilon_j\) of individual performance, where \(\epsilon_j\) is idiosyncratic noise. What drives the above results is that the variance of the shock is increasing in employment so that effort becomes harder to measure as the firm grows. As long as \(\epsilon_j\) has that property, we should expect similar results. For example, if \(\epsilon_j\) is i.i.d. normal with mean zero and variance \(\sigma_j^2L_i\) then clearly our results would be completely unaffected.

### 4 The Long Run

In the long run, firms are free to adjust their capital stocks. The problem of firm \(i\) is to choose capital \(K_i \geq 0\) and labor \(L_i \geq 0\) to maximize expected profit (23). The long-run costs and benefits of capital, given that employment is chosen optimally, are evident from (23). The benefit is that

\(^{19}\)Note that Garen uses this measure of industry risk to proxy individual firm risk. After adding industry dummies, the coefficient on industry risk becomes insignificant. According to Garen (p. 1189), “This is not surprising given that much of the variation in industry R&D is explained by aggregate industry dummies.”
capital increases expected output as can be seen in the numerator of the first term. The direct cost of capital is the final term $\rho_i K_i$. Since an increase in capital induces an increase in employment, the indirect costs of capital are the increase in the risk premium (the denominator of the first term) as well as the external cost of employment $\bar{\pi} L_i$. The optimal capital stock balances these tradeoffs. Unfortunately, the firm’s maximization problem does not have a closed-form solution for any standard functional form for $f(K_i)$. We can, however, use lattice programming methods to achieve the main objective of the paper, which is to obtain comparative statics results that can be taken to the data.

**Proposition 7** Given $r$ and $\bar{\pi}$, assume there exists a subset $T^l$ of parameter space

$$P^l_i = (p_{M_i}, A_i, 1/\sigma^2_i, 1/\rho_i)$$

where the problem of firm $i$ has a unique interior solution $K_i(P^l_i)$ and $L_i(P^l_i)$ and expected profit is nonnegative. Then on this region $T^l$ of parameter space, $K_i$ and $L_i$ are nondecreasing in each of the parameters $P^l_i$.

The proof is in the appendix. These comparative statics results are intuitively straightforward given that capital and labor are complements in the production function (3), so a change in some parameter which increases one will tend to increase the other. An increase in the price $p_{M_i}$ or total factor productivity $A_i$ makes employment and capital more valuable to the firm which leads to an increase in both. An increase in the marginal variance $\sigma^2_i$ or the CARA coefficient $r$ increases the internal cost of employment in terms of the risk premium. This will reduce employment and therefore capital. Finally, an increase in the cost of capital $\rho_i$ will lead to a reduction in capital in the first instance followed by a complementary reduction in employment.

The more interesting comparative statics results pertain to incentives and expected wages and profits. Note that the above expressions for employment (24), incentives (27), and the expected wage (32) remain valid in the long run. For these endogenous variables we have strict comparative statics results instead of the weak ones in Proposition 7 above. To simplify the statements of the next two results, we assume the comparative statics results for capital in the above proposition are strict instead of weak. For example, we assume $K_i$ is increasing rather than nondecreasing in $A_i$.

**Assumption 2** Assume the comparative statics results for capital $K_i$ in Proposition 7 are strict instead of weak.

To obtain the reciprocal of the long-run incentive, we substitute the optimal capital stock $K_i(P^l_i)$
into the expression (30) for the reciprocal of the short-run incentive

\[ \tilde{\beta}_i(P_l^i) = \frac{A_i f(K_i(P_l^i))}{\sqrt{2\pi}}. \] (48)

Similarly, the long-run expected wage \( w_i(P_l^i) \) is obtained by substituting \( K_i(P_l^i) \) into the short-run expected wage (32) and long-run expected profit \( \Pi_i(P_l^i) \) by substituting \( K_i(P_l^i) \) and \( L_i(P_l^i) \) into short-run expected profit (23). Already we observe some important differences between the short and long runs. First, the long-run incentive \( \beta_i \) now depends directly on the marginal variance \( \sigma_i^2 \) and the CARA coefficient \( r \) whereas the short-run incentive did not. This is because capital is now endogenous and depends on those parameters. Furthermore, all the endogenous variables now depend on the cost of capital \( \rho_i \).

**Proposition 8** The long-run expected wage \( w_i(P_l^i) \), expected profit \( \Pi_i(P_l^i) \), and reciprocal of the incentive \( \tilde{\beta}_i(P_l^i) \), are all increasing in each of the parameters \( P_l^i \) on the region \( T^i \).

According to this result, an increase in the CARA coefficient \( r \) or the marginal variance \( \sigma_i^2 \) leads to stronger incentives \( \beta_i \) and a lower expected wage \( w_i \). In the short run there is no direct relationship between incentives and risk whereas in the long run there is a direct and positive relationship. The long-run result may seem counterintuitive, especially from a contract theory perspective grounded in the standard risk-reward tradeoff, but as we have seen there is substantial evidence to support this prediction.

Intuitively, an increase in \( r \) or \( \sigma_i^2 \) with capital held fixed increases the cost of incentives in terms of the risk premium but also reduces employment which makes effort easier to measure. As we saw in the previous section, these two effects cancel and there is no effect on the optimal incentive in the short run. But in the long run capital is endogenous and an increase in \( r \) or \( \sigma_i^2 \) reduces capital which makes effort less valuable to the firm and reduces the sensitivity of effort to incentives. Both of these effects point towards weaker incentives. But the reduction in capital also leads to a further complementary decline in employment which again makes effort easier to measure. The second effect dominates, so the optimal incentive is increasing in \( r \) and \( \sigma_i^2 \) in the long run when capital is endogenous. Finally, the reduction in employment lowers the risk premium which implies a smaller compensating differential for risk. The expected wage \( w_i \) therefore declines.

The same logic explains the effects of increases in the price \( p_{M_i} \), total factor productivity \( A_i \), and the cost of capital \( \rho_i \). The results for \( \rho_i \) are significant in that we are unaware of any other theoretical model or empirical evidence which draws a connection between incentive intensity and the cost of capital. In our model, an increase in \( \rho_i \) leads to a reduction in not only capital but also
employment, which makes effort easier to measure and leads to stronger incentives.

Under the assumption that the vector $P^l_i$ of exogenous parameters in (47) is associated, the model generates the following long-run empirical predictions.

**Proposition 9** If $P^l_i$ is associated on $T^l$ then the vector

$$\left( \tilde{\beta}_i(P^l_i), K_i(P^l_i), L_i(P^l_i), w_i(P^l_i), \Pi_i(P^l_i), P^l_i \right)$$

is associated on $T^l$ for any given $r$ and $\Pi$.

We first consider the appropriateness of the assumption that $P^l_i$ is associated. We focus on the new parameter $\rho_i$, which reflects the cost of borrowing for firm $i$. The inverse $1/\rho_i$ should be positively correlated with the price $p_M$, total factor productivity $A_i$, and the inverse $1/\sigma^2_i$ of the marginal variance because all of these parameters increase expected profit and therefore reduce credit risk.

This result strengthens and extends the corresponding short-run result in the previous section. It strengthens our previous results in the sense that we no longer need to assume affiliation but rather the weaker assumption of association. In practical terms, it means that we no longer have to control for capital stocks in the relevant regressions. The result extends previous ones in the sense that most of the short-run predictions in the previous section are preserved in the long run. For example, incentives continue to be negatively correlated with firm size as measured either by employment or capital and the expected wage remains positively correlated with both firm size and expected profit. Note that Lemmas 1 and 2 continue to hold in the long run, so the covariance between wages $I_i$ and the other exogenous and endogenous variables will be positive under the same assumptions as before. The exceptions stem from the fact that capital is now endogenous and the cost of capital $\rho_i$ now influences the endogenous variables. In particular, the long run incentive is positively correlated with the marginal variance $\sigma^2_i$ and the endogenous variables are all negatively correlated with the cost of capital $\rho_i$.

## 5 Conclusion

In this paper we extended the workhorse moral hazard model in Holmström and Milgrom (1987) to encompass not only incentives but also employment, capital, and the external labor market. The resulting model is simple, with straightforward explanations for sometimes counterintuitive results, and has the same wide scope of application as the textbook neoclassical model. It therefore has the
potential to serve as a benchmark model for thinking and making predictions about employment, capital, incentives, wages, and profit. The main assumptions of the model are that effort becomes harder to measure as the firm gets larger and the exogenous parameters are affiliated. We showed that the endogenous variables and exogenous parameters can be defined in such a way that the former are all at least nondecreasing in the latter. An application of Holmström and Milgrom (1994) allowed us to conclude that the endogenous and exogenous variables should exhibit nonnegative pairwise covariances in data sets with the appropriate time frame.

The potential of the model to serve as a useful theoretical benchmark is illustrated by its ability to explain several stylized facts: why incentives decline but wages rise with firm size as measured by either employment or capital, the mixed nature of the evidence on the risk-reward tradeoff, and the positive correlation between wages and profits. These predictions are straightforward consequences of the fact that employment is endogenous and effort becomes harder to measure as employment grows. In our model, the firm under moral hazard operates within a competitive labor market which determines the overall payoff of the workers, while incentives and wages are set within the firm subject to the endogenous participation constraint. This aspect of the model allows us to make novel predictions about the relationship between incentives and aggregate labor market conditions. In particular, any labor market phenomenon which increases the workers’ total payoff (e.g., a decrease in labor force participation) will lead to stronger incentives. When capital is endogenous, we obtain the counterintuitive but empirically relevant prediction that incentives are positively related to risk in the long run. Another novel prediction is that incentives should be positively related to the cost of capital, which makes a potential connection between incentive contracting and macroeconomic conditions in the form of interest rates.

6 Appendix

Proof of Proposition 7. We apply Theorem 2.3 in Vives (1999, p. 26). We drop the subscript \( i \) throughout the proof. Let \( X \) be the set of all \((K, L) \in (0, \infty) \times (0, \infty)\). Note that \( X \) is a lattice and \( T^i \) is a partially ordered subset of \( \mathbb{R}^4 \) which is nonempty by assumption. We denote derivatives using subscripts. The following crosspartials show that expected profit (23) is strictly
supermodular on $X$ with strictly increasing differences on $X \times T$.

$$\Pi_{KL} = \frac{A^6 p_M^2 f(K)^5 f'(K)}{[A^2 f(K)^2 + L r \sigma^2]^3} \left[A^2 f(K)^2 + 3 L r \sigma^2\right] > 0$$

$$\Pi_{KA} = \frac{2 A^3 L p_M^2 f(K)^3 f'(K)}{[A^2 f(K)^2 + L r \sigma^2]^3} \left[A^4 f(K)^4 + 3 A^2 L r \sigma^2 f(K)^2 + 4 L^2 r^2 \sigma^4\right] > 0$$

$$\Pi_{KPM} = \frac{2 L p_M f(K)}{[A^2 f(K)^2 + L r \sigma^2]^2} \left[A^6 f(K)^5 + 2 A^4 L r \sigma^2 f(K)^3\right] > 0$$

$$\Pi_{K \sigma^2} = -\frac{2 A^4 L^3 p_M^2 \sigma^2 f(K)^3 f'(K)}{[A^2 f(K)^2 + L r \sigma^2]^3} < 0$$

$$\Pi_{LA} = \frac{A^5 p_M^2 f(K)^6}{[A^2 f(K)^2 + L r \sigma^2]^3} \left[A^2 f(K)^2 + 3 L r \sigma^2\right] > 0$$

$$\Pi_{LPM} = \frac{A^6 p_M f(K)^6}{[A^2 f(K)^2 + L r \sigma^2]^2} > 0$$

$$\Pi_{L \sigma^2} = -\frac{A^6 L p_M^2 f(K)^6}{[A^2 f(K)^2 + L r \sigma^2]^3} < 0$$

$\Pi_{L \rho} = 0$, and $\Pi_{K \rho} = -1$. From Theorem 2.3 cited above we conclude that $K$ and $L$ are at least weakly increasing in the parameters.

7 References


