Organizational Design and Technology Choice under Intrafirm Bargaining

By Lars A. Stole and Jeffrey Zwiebel*

We consider a wide number of applications of an intrafirm bargaining game within organizations where employees and the firm engage in wage negotiations. Under our presumption that contracts cannot bind employees to the organization, the resulting stable wage and profit profiles give rise to an objective function for the firm that places weight on inframarginal profits in an economically significant manner. We in turn employ this methodology to explore applications of organizational design, hiring and capital decisions, training and cross-training, the importance of labor and asset specificity, managerial hierarchies, preferences for unionization, responses to competition, and internal capital budgeting. (JEL C70, D23, G31, J30, L20.)

Labor contracts are frequently incomplete, with only a limited capability to bind either party to the relationship. Employees are generally free to quit at will, and firms typically can dismiss part of their labor force subject to only limited penalties or costs.1 Given benefits from joint production, we expect firms with productive capital to hire labor, but the ability of any employee or the firm to reenter into wage negotiations at any time prior to production places particular restrictions on the equilibrium wage profile. Focusing on the firm’s optimal choices in such a bargaining framework, we consider a wide range of economic applications regarding labor decisions, technological choice, and organizational design using a novel bargaining methodology.

1 See, for examplt, Richard Posner (1986 p. 306), who states, “employment at will is the usual form of [the] labor contract.” While there are some notable limitations under which courts have held employees responsible for damages for violating long-term labor obligations, such examples in practice appear to be quite specialized. Any attempt at widespread application of such long-term contractual binding of labor is likely to be ruled in conflict with restrictions on involuntary servitude. Thus, both the contract-theory and the managerial-incentives literatures frequently take long-term labor contracts to be nonbinding on the employee, as we do here.

Somewhat more problematic is our presumption that labor relations are nonbinding on the firm as well. In practice, there are clearly a number of restrictions and costs that firms are likely to face in terminating employees; the extent of such restrictions in many European countries has in fact been blamed by some for high unemployment levels. However, under standard weak assumptions that we impose on production functions, such a constraint on employee terminations will be irrelevant for almost all of our analysis: our results would be identical if it were only on the employees that labor contracts are nonbinding.
Following Ronald Coase's (1937) seminal work on the nature of the firm, economists have spent much time debating and defining the boundaries of the firm and the marketplace. Most recently, the papers of Sanford Grossman and Oliver Hart (1986) and Hart and John Moore (1990) ("GHM") have refined the idea of the firm as essentially an allocation of property rights and residual rights of control over assets. Fundamental to the GHM "property rights" theory of the firm is that contracts are incomplete, and ex post bargaining generally results in return streams that are suboptimal for specific investments from an ex ante view point. This line of research, however, has generally focused on the question of what determines the boundaries of a firm and how assets should optimally be allocated, taking as given the underlying asset structure, the number of productive agents, and their organizational relationships to one another. We seek an entirely different approach, asking questions regarding the optimal nature of assets, the number of productive agents, and the relationship of labor to capital in the optimal organizational design.

Our approach is similar in some manners to that of Michael Jensen and William Meckling (1976) who take the firm as a nexus of contracts and lay out the program of studying the optimality of various underlying contractual arrangements and the outcome of the complex equilibrium process within the firm. We take as a working definition that a firm is a set of productive assets, which produces output that depends upon the allocation of laborers to those assets. To bring the analysis into clear focus, we assume that the firm's assets are owned by a single individual. Consequently, we put aside the question of optimal asset ownership and, rather, devote our attention to other issues. To this extent, we do not address the issue of a firm's external boundary, but rather a formal theory of the firm's internal bargaining process and its economic ramifications.

Specifically, we take labor contracts to be nonbinding in nature. The focus of our paper is on situations in which employees have the ability to bargain directly with the firm either because of their importance within the firm (managers, division heads, key engineers, etc.) or the small size of the firm. We model labor contracts as an agreement for a wage which the firm (here, the owner of the productive capital) pays the employee, provided that the employee actually carries out the contracted productive services. At any time before production takes place, an employee may approach the firm and enter into wage negotiations. Likewise, the firm may choose at any moment before production to call the employee in for wage negotiations. Hence, with labor contracts, the nature of the contractual incompleteness is the inability of either party to commit to a future wage and employment decision. In this sense, our stylized firm is quite similar to a firm with "employment-at-will" contracts. Although our model incorporates the additional feature of labor holdup, we will frequently refer to our firm as an "employment-at-will" (or simply "at-will") firm to distinguish it from the neoclassical complete-contract benchmark. Given our employment-at-will contractual incompleteness, we consider wage contracts which are immune to renegotiations in the intrafirm negotiation environment.

Implicitly, we also assume that either (i) the firm cannot commit (i.e., write a credible contract) to fix a technology and hire an agreed-upon number of workers and not subsequently hire more or (ii) the employees are financially constrained so as to be unable to pay an entry fee to the firm. When neither assumption is true, the first best can be obtained by employees paying an entry fee equal to the present value of their excess bargaining rents and the firm, in turn, committing to hire at the first-best level using the first-best technology and labor-allocation choices.

Several papers in the literature of incomplete contracts have considered the ability of negotiating parties to obtain ex ante efficiency in light of the holdup problem. Most notably, see Hart and Moore (1988), Tae-Yeong Chung (1991), William Rogerson (1992), Bentley MacLeod and James Malcolmson (1993), Philippe Aghion et al. (1994), Georg Nölkeke and Klaus Schmidt (1995), and Aaron Edlin and Stefan Reichelstein (1995). Unlike these papers, which typically obtain the first best by utilizing a limited set of contractual variables, we do not allow

---

In order to determine the economic consequences of contractual incompleteness and intrafirm bargaining, we need to characterize the set of reasonable outcomes from an internal wage-bargaining process. Elsewhere, in Stole and Zwiebel (1996), we have studied a general class of bargaining games and their resulting equilibrium outcomes. Although all of the qualitative results in this paper hold across this general class, here we focus on the simplest bargaining process in which workers and the firm have equal bargaining power. In Section I, we present the simple bargaining environment and discuss the nature of stable wage and profit profiles. Using these results we characterize the firm's reduced-form profit function and its optimal choices in a variety of settings.

In Section II, we apply our bargaining results to economic applications of input choice. Specifically, we study the comparative statics of labor and capital choice when the firm is subject to internal wage negotiations. We compare these results for our employment-at-will firm to that of the standard neoclassical paradigm. We find striking implications of overemployment and capital underutilization. Throughout our results, we find that, unlike its neoclassical counterpart, the employment-at-will firm generally cares about inframarginal profits which are not reached in equilibrium. To shed light on this phenomena, in Section III we investigate the effect of inframarginal changes in various economic settings by examining the firm's front-load factor—the extent to which margins are loaded up front (i.e., are realized with low levels of labor). We find that such a factor can be used to explain many diverse economic questions. In particular, we apply the analysis to the question of labor specificity, the preference for unionization, and the effect of competition on observed firm productivity.

In Section IV, we consider problems of organization. Suppose that the firm can choose two productive assets or groups, A and B, each of which has an exclusive assignment of employees chosen by the firm. How will a firm tailor its asset choices to one another in light of nonbinding labor contracts? Specifically, are there intraorganizational misallocations across groups? We generally find that the answer is affirmative. We consider, in turn, three applications to illustrate the nature of such distortions. First, we examine the effect of a two-tier hierarchy on a firm's employment decision. The ability of the higher tier to hold up more productive value of the firm may possibly lead to top-heavy hierarchies emerging in employment-at-will firms. We then consider the internal allocation of capital in a capital-budgeting framework. Here we find that the owner of the firm does not allocate capital according to marginal profitability, but instead allocates on the basis of average inframarginal productivity, leading to pronounced departures from the neoclassical optimum. Finally, we consider the value to the firm from cross-training its workers. The flexibility afforded by cross-training leads to an increase in the at-will firm's returns, even though workers are not reallocated in equilibrium. This suggests that, when cross-training entails a cost, a socially excessive amount of cross-training occurs in order to hold down labor rents.

Section V concludes by suggesting further applications and extensions using the present methodology.

I. Theoretical Foundations

The bargaining game we consider for wage determination within a firm is based on that in Stole and Zwiebel (1996), in which we analyze a general bargaining environment, discussing in detail the properties of our solution and various related theoretical issues. Presently, we confine ourselves to a particularly simple version of this game, in which we presume that bargaining power is identical between the employee and the firm. However, we will briefly mention how the applications in this paper are robust to the more general bargaining environment. Here, we simply state results together with some intuition; all omitted proofs of results in Section I, together with a more rigorous theoretical treatment of the bargaining game, are contained in Stole and Zwiebel (1996).
We consider a noncooperative dynamic bargaining game, in which a firm, endowed with a production function, hires labor and other factors of production. We will consider both the case of homogeneous labor and that of differentiated labor, which we can alternatively interpret as an assignment of employees to different productive tasks or groups. Presently, we give an intuitive description of our game, starting with the case of homogeneous labor.5

In line with our notion of renegotiation-proof contracts, we look for what we call a "stable" wage and profit profile. By stable, we mean a profile in which, prior to production, no individual employee can benefit from renegotiating wages with the firm, and the firm cannot benefit from a renegotiation with any employee given the further wage renegotiations that such a renegotiation would induce. In all such negotiations, we presume that the employee and the firm evenly split the joint surplus from their relationship relative to their respective outside options. For the employee, this outside option is simply the reservation wage w, while for the firm it is the outcome of such a bargaining process with one less employee in the firm. As such, it should be clear that the outcome to our game is built up iteratively. In order to characterize a stable profile for a firm with n employees, we must first characterize such an outcome for n – 1 employees to determine the outside option for the firm. We assume that the firm cannot replace an employee who is fired or who quits prior to production; this is precisely the nature of employee holdup power in our model.

We will let F(n) represent the firm's revenue or output as a function of the number of employees; later when we consider the choice of other inputs (e.g., capital) we will use F(n, x). We use tildes to denote the choices and outcomes of our employment-at-will firm (in contrast to the neoclassical firm); as such, we let \( \tilde{w}(n) \) and \( \tilde{\pi}(n) \) represent the stable wage and profit profile to our game when there are n employees. Because all surplus not paid to employees is retained by the firm, \( \tilde{\pi}(n) = F(n) - \tilde{w}(n)n \). Lastly, we will use \( \pi(n) = F(n) - wn \) to represent the profits of a neoclassical firm that can utilize labor at the outside option of w. We assume that this neoclassical profit function is quasi-concave. It is worth noting here that the outcomes of the at-will setting can be shown to converge to those of the neoclassical setting in the limit as all bargaining power in employee-firm negotiations is given to the firm.

First consider the employment-at-will firm with one employee. In bargaining between the firm and the employee, the outside options of the two parties upon failing to reach an agreement are given by \( F(0) \) and w, respectively. Consequently, an even division of the surplus implies that

\[
\Delta F(1) - \tilde{w}(1) = \tilde{w}(1) - w
\]

or \( \tilde{w}(1) = \frac{1}{2}[\Delta F(1) + w] \), where \( \Delta \) is the first-difference operator.

Now consider the same firm with two employees. While each employee's outside option is still w, the firm's outside option is now given by the equilibrium profit that would be obtained from production with one remaining employee if current negotiations break down. Specifically, if an employee leaves (i.e., is fired or quits), the remaining employee and the firm would renegotiate wages to \( \tilde{w}(1) \) (i.e., the one-employee outcome). Consequently, splitting the surplus in negotiations when there are two employees implies that

\[
\Delta F(2) - \tilde{w}(2) + [\tilde{w}(1) - \tilde{w}(2)] = \tilde{w}(2) - w.
\]

The left-hand side of this expression is simply the difference in firm profits between employing two workers and one, \( \tilde{\pi}(2) - \tilde{\pi}(1) \), given by the second margin of production net the difference in equilibrium wages paid. Substituting in the result for \( \tilde{w}(1) \) above yields

\[
\tilde{w}(2) = \frac{1}{2}[\Delta F(2) + \tilde{w}(1) + w]
\]

\[
= \frac{1}{2} \Delta F(2) + \frac{1}{6} \Delta F(1) + \frac{1}{2} w.
\]

---

5 For a more rigorous exposition, including an extensive-form representation of the game in which our outcome is the unique subgame-perfect equilibrium, we refer the reader to Stele and Zwiebel (1996).
By induction, for any \( n \), it then follows that

\[
\Delta F(n) - \hat{w}(n) + \{ (n - 1) \\
\times [\hat{w}(n - 1) - \hat{w}(n)] \} = \hat{w}(n) - w
\]

from which we obtain the difference equation,

\[
(1) \quad \hat{w}(n) = \frac{1}{n + 1} \\
\times [\Delta F(n) + (n - 1)\hat{w}(n - 1) + w].
\]

This difference equation has a unique solution. Providing that the path of wages satisfies the workers’ outside option, \( \hat{w}(i) \geq w \), \( \forall i \leq n \)—which we refer to as feasibility— this solution is also the equilibrium outcome of the bargaining process. The following result characterizes this outcome.

**Result 1:** If there are \( n \) employees in the firm and the solution to (1) is feasible, then the stable wage and profit profiles are given by

\[
(2) \quad \hat{w}(n) = \frac{1}{n(n + 1)} \sum_{i=0}^{n} i\Delta F(i) + \frac{1}{2} w \\
(3) \quad \bar{\pi}(n) = \frac{1}{n + 1} \sum_{i=0}^{n} \pi(i).
\]

Equation (2) follows directly from solving (1); equation (3) is determined by substituting (2) into \( \bar{\pi}(n) = F(n) - \hat{w}(n) n \), summing by parts, and then using the definition of neoclassical profits to simplify the resulting expression.

There are several noteworthy features of this characterization. Equation (2) indicates that the equilibrium wage depends upon a weighted average of marginal products, with increasing weight in the wage expression the closer to marginal the product. Thus, the more up-front a margin of production, the smaller a share of that margin an employee obtains, and the more profit the firm receives. Such a notion of “front-loaded” margins is exploited in the applications that follow. Equation (3) is also quite notable; it states that the at-will firm’s profits with \( n \) employees can be expressed concisely as the average on the neoclassical firm’s profits from 0 to \( n \) employees. This result, rather striking in its simplicity, will be employed throughout our applications.

It is also worth mentioning that this result retains the features that are important to our economic applications under the more general bargaining game considered in Stole and Zwiebel (1996). Both the monotonicity of marginal weights in the equilibrium wages in (2) and the averaging of the neoclassical profits to obtain at-will profits in (3),\(^6\) remain for any exogenously specified bargaining power, including those allowed to vary with number of employees. Because it is precisely these properties upon which our applications depend, the results that we present should be understood to be more generally applicable than for the specific framework we consider. Furthermore, it is notable that the payoffs given in Result 1 are equivalent to the standard cooperative-game solution concept of Shapley values for a corresponding cooperative game. As such, we can alternatively give our bargaining game a cooperative interpretation, whereby the employees’ and the firm’s payoffs depend on their strategic importance in coalition formation.

In a number of applications to follow, it will be useful to consider the outcome under continuous labor allocations. To this end, we can either take the limit to our solution as bargaining units become small or solve the differential equation that is the limiting case of our difference equation. Either approach yields Result 2.

**Result 2:** Suppose that there are \( n \) units of infinitely divisible labor and the solution to the differential equation that is the limiting case of (1) is feasible. Then,

\[
(4) \quad \bar{\pi}(n) = \frac{1}{n} \int_{0}^{n} \pi(s) \, ds.
\]

We will also consider a number of applications in which the firm has an input of several types of differentiated labor. In many important economic settings, the assumption

\(^6\) In the general case, \( \bar{\pi} \) can be expressed as a weighted average of \( \pi \) from 0 to \( n \), where weights depend on “marginal bargaining power.”
that labor within a firm is interchangeable is particularly restrictive. A typical firm may consist of hundreds of different positions, with very little opportunity for perfect substitution across employees. When a senior manager of marketing threatens to leave the firm, for example, it is difficult for the firm’s owners to argue credibly that they will replace her with an equally efficient senior manager from production. To this end, we need a generalization of the bargaining environment that explicitly recognizes that not all labor is identical. We accomplish this by considering different categories of labor; although labor within a category is homogeneous, across categories labor may differ arbitrarily in productive uses and outside wages. Note that such a representation of heterogeneity also allows the possibility that all workers are different, and therefore the number of categories equals the number of employees. Letting \( M \) denote the number of different categories or “assets” with which to assign labor, we consider the production function of the form \( F(n) \), where \( n = (n_1, \ldots, n_M) \).

Additionally, allowing for the possibility of differing reservation wages, we take \( w = (w_1, \ldots, w_M) \) to denote the vector of reservation wages over the \( M \) assets. In this setting, we obtain the following result using arguments analogous to the single-setting argument.

**Result 3**: Suppose that the labor–asset allocation is given by the vector \( n \), and that labor is finite. If the solution to the partial difference equation system corresponding to the bargaining game is feasible, then the equilibrium to the noncooperative game has profit given by

\[
\bar{\pi}(n) = \frac{1}{N+1} \sum_{i_1=0}^{n_1} \cdots \sum_{i_M=0}^{n_M} \left( \frac{n_1}{i_1} \right) \cdots \left( \frac{n_M}{i_M} \right) \frac{N}{\Sigma i_j} \pi(i_1, \ldots, i_M)
\]

where \( N = \sum_{m=1}^{M} n_m \). If instead labor is infinitely divisible, then the equilibrium to our bargaining game has profits given by

\[
\bar{\pi}(n) = \frac{1}{N} \int_0^N \{ F(s\alpha) - s\alpha \cdot w' \} \, ds
\]

where \( \alpha = (\alpha_1, \ldots, \alpha_M) \) and \( \alpha_i = n_i/N \).

This result indicates that, once again, at-will firm profits are given by an average of neoclassical profits for inframarginal labor configurations. In the continuous case, this average is simply that of neoclassical profits along the vector from 0 to \( \alpha N \) employees. Additionally, we have shown elsewhere that both characterizations in Result 3 represent the Shapley value of the corresponding cooperative bargaining game; in the case of the continuous labor setting, this is commonly referred to as the Aumann-Shapley (1974) value.

We now state several results characterizing the firm’s optimal labor and input decisions. In order to explore how bargaining affects input choice, we augment our production function with an argument, \( x \), representing a nonlabor input. In many of this paper’s applications, \( x \) can be thought to represent capital choice, at the cost of \( r \) per unit. (Results can easily be extended to the case where \( x \) is a multidimensional vector of inputs.) We generally take capital to be fixed at the time of bargaining, though we also discuss the case of adjustable capital in Section II.

We will denote the optimal labor and input decisions for the neoclassical firm by \( n^* \) and \( x^* \), respectively, and for the at-will firm we will denote these decisions by \( \tilde{n}^* \) and \( \tilde{x}^* \). In order to state these results precisely in the discrete labor setting, it is useful to define the relationship \( y(n) \doteq z(n) \) to indicate that \( y(n) - z(n) > 0 \simeq y(n + 1) - z(n + 1) \); we say that \( y \) equals \( z \) at \( n \) to within integer rounding over its argument.\(^7\) Given the results of our bargaining game, the following pair of results characterize the optimal labor and factor input decisions by the at-will firm in a simple and intuitive manner and indicate that the at-will firm overhires labor to the point where wages are driven down to the outside option.

\(^7\)Intuitively, if \( y \) and \( z \) were continuously and monotonically extended to \( \bar{y} \) and \( \bar{z} \) over the interval \([n, n+1]\), \( y(n) \doteq z(n) \) or \( z(n) = y(n) \) if and only if \( \bar{y}(m) = \bar{z}(m) \) for some \( m \in [n, n + 1] \).
Result 4: Suppose that \( \pi(n, x) \) is quasiconcave and \( \tilde{\pi}(n, x) \) has an interior optimum over \( (n, x) \):

1. When labor is discrete, \( \{\tilde{n}^*, \tilde{x}^*\} \in \text{arg} \max_{n, x} \tilde{\pi}(n, x) \) satisfies

\[
\pi(\tilde{n}^*, \tilde{x}^*) = \tilde{\pi}(\tilde{n}^*, \tilde{x}^*)
\]

\[
\sum_{i=0}^{\tilde{n}^*} \frac{\partial \pi(i, \tilde{x}^*)}{\partial i} = 0.
\]

2. When labor is continuous, \( \{\tilde{n}^*, \tilde{x}^*\} \in \text{arg} \max_{n, x} \tilde{\pi}(n, x) \) satisfies

\[
\pi(\tilde{n}^*, \tilde{x}^*) = \tilde{\pi}(\tilde{n}^*, \tilde{x}^*)
\]

\[
\int_{0}^{\tilde{n}^*} \frac{\partial \pi(s, \tilde{x}^*)}{\partial s} \, ds = 0.
\]

Result 5: Optimally, the at-will firm chooses to overemploy: \( \tilde{n}^* > n^* \). Furthermore, at this optimal level of employment \( \tilde{n}^* \), the internal wage is bid down to the outside option: \( \tilde{\omega}(\tilde{n}^*) = \omega \) for the discrete labor setting or \( \tilde{\omega}(\tilde{n}^*) = \omega \) for the continuous labor setting.

The two conditions of Result 4 are simply the first-order conditions for \( n \) and \( x \), respectively; the assumptions made for the neoclassical profit function \( \pi \) ensure feasibility and an interior solution to the at-will profit function \( \tilde{\pi} \).

This result, while quite powerful, is easy to understand in light of our characterization of negotiated profits in equations (3) and (4). Intuitively, at-will profits at any level of employment \( n \) are given by the average neoclassical profits from \( 0 \) to \( n \). But such an “average” function obtains its maximum at precisely the point where the “marginal” function (here, the neoclassical profit function) cuts it from above. (Save for a sign change, this is no different from the basic microeconomic result that a firm’s marginal-cost curve intersects its average-cost curve from below at the latter’s minimum.) Thus at \( \tilde{n}^* \), the neoclassical and the at-will profits must be equal (to within integer rounding on \( n \) in the discrete case). And since the marginal curve cuts the average curve from above, the first half of Result 5 follows: \( \tilde{n}^* > n^* \) and overemployment results.

Furthermore, equal profits at \( \tilde{n}^* \) in turn imply that wages must be equal as well (the second half of Result 5), since in both cases profits are given by the same production net total wages paid. By definition, neoclassical wages are always \( \omega \); thus, the at-will firm hires until wages are driven down to this outside option as well. Therefore, while the ability of employees to bargain in the at-will firm yields higher wages for any given number of workers \( n \), this bargaining is exactly offset by the choice of this firm to overhire to \( \tilde{n}^* \), leading to equilibrium wages that are the same across both types of firms. This relationship in turn readily allows for the embedding of our partial analysis in a general-equilibrium model with both neoclassical and at-will firms. While most of this paper’s applications ask questions about firm decisions at this equilibrium configuration, we consider several examples that address wage-negotiation distortions holding the number of workers fixed, which consequently give rise to a wage premium for employees of the at-will firm. Additionally, the applications of hiring costs and adjustment dynamics to exogenous shocks considered in Section II generate a wage differential between the at-will and neoclassical employees (made up for \textit{ex ante} by the imposition of a hiring cost).

Finally, the first-order condition for input \( x \) in Result 4 indicates that the marginal returns to \( x \) obtained by the at-will firm are the neoclassical marginal returns averaged over \( 0 \) to \( n \) employees. Intuitively, a firm’s ability to capture the marginal value added by additional capital \( \Delta x \) depends on the extent to which the holdup ability of labor affects this margin (i.e., on how the productivity of this margin changes with the number of employees). Consequently, if capital and labor are complements (substitutes), at a given level of \( (n, x) \), the marginal benefits of capital are lower (higher) for the at-will firm than for the neoclassical firm.

* While our equilibrium analysis in this paper is almost exclusively partial, many of the applications could be extended in an interesting way to a general-equilibrium framework. Such an extension would allow one to explore how the two different firms respond to different exogenous shocks in equilibrium.
It turns out that the distortion can be conveniently characterized with a single statistic that we call the front-load factor. We define this statistic, for the discrete and continuous cases, respectively, by

$$\gamma(F, n) = 1 - \frac{1}{\pi(n)} \sum_{i=0}^{n} \frac{i}{n + 1} \Delta \pi(i)$$

and

$$\gamma(F, n) = 1 - \frac{1}{\pi(n)} \int_{0}^{s} \frac{\pi'(s)}{n} ds.$$  

Intuitively, this statistic measures to what extent neoclassical profit margins (and thus, equivalently, production margins) are realized through earlier rather than later units of production. Thus, the later the unit of production, the greater the negative coefficient (e.g., the $i$th unit, in the discrete setting, is weighted by $-i/[n + 1]$). The specific normalization we have chosen results in $\gamma \in [0, 1]$, whenever \( \tilde{\omega} \) is feasible, and the following concise result.

**Result 6:** In both the discrete and continuous settings where \( \tilde{\omega}(n) \) is feasible,

$$\hat{\pi}(n) = \pi(n) \gamma(F, n).$$

This follows directly from integrating (or, for the discrete case, summing) $\tilde{\pi}(n)$ by parts. It indicates immediately both that firms prefer technologies with higher front-loadings, and that any distortion between the neoclassical input choice and the at-will firm’s input choice emerges exclusively from the effect of the choice variable on $\gamma$. In addition, combining this result with Result 4 (i.e., $\pi(\tilde{\pi}^*) = \hat{\pi}(\tilde{\pi}^*)$) implies that at-will firms hire up to the point where the front-load factor is increased to 1 (within integer roundings for the discrete labor setting).

A useful set of sufficient conditions for one technology, $\pi^1$, to be relatively more front-loaded at $n$ than another, $\pi^2$, is that the technologies are equally efficient at $n$ [i.e., $\pi^1(n) = \pi^2(n)$] and $\pi^1(s) \geq \pi^2(s)$ for all $s \in [0, n]$, with strict inequality for some measurable subset. Thus, for example, if $\pi^1(0) = \pi^2(0)$ and $\pi^1(n) = \pi^2(n)$ and if $\pi^1$ is a concave transformation of $\pi^2$ over the range $[0, n]$, then $\pi^1$ is more front-loaded relative to $\pi^2$ at $n$. Note, however, that the converse is not true. Front-loading is a far weaker notion than concavity, inducing a complete order over profit functions (at a given level of labor), rather than only the partial ordering of concavity.

Finally, it is worth noting that these optimal-input-choice results hold for the multiasset environment as well. That is, the multiasset at-will firm chooses both the total amount of labor $N = \sum_{i=1}^{M} n_i$ and the allocation of labor across assets $\alpha = (\alpha_1, ..., \alpha_M)$ in a manner that equates neoclassical and at-will profits, and consequently such a firm drives the wage of employees on each asset down to their reservation value. We state this result for the simpler continuous labor environment; its proof follows from the first-order conditions for the continuous multi-asset $\hat{\pi}$ in Result 3.

**Result 7:** Suppose that $\pi(\alpha N)$ is quasiconcave and $\hat{\pi}(\alpha N)$ has an interior optimum over $(\alpha, N)$: then the optimal $\{\alpha^*, N^*\}$ for the at-will firm satisfies

$$\pi(\alpha^*_N N^*, ..., \alpha^*_M N^*) = \hat{\pi}(\alpha^*_N N^*, ..., \alpha^*_M N^*)$$

and

$$\int_{0}^{N^*} s \frac{\partial \pi}{\partial n_j} (\alpha^*_N, ..., \alpha^*_M) ds = \int_{0}^{N^*} s \frac{\partial \pi}{\partial n_k} (\alpha^*_N, ..., \alpha^*_M) ds$$

$$= 1 \quad \forall j, k = 1, 2, ..., M.$$  

It is straightforward to show that, in the presence of input choice, as in the single-asset setting, inputs would be chosen so that their marginal contribution, averaged appropriately over all labor configurations, is zero. Note that this result implies that at this point, (partial) front-load factors defined for individual assets must equal 1 across all assets. As such, the multiasset setting is similar to the single-asset case for the choice of firm scale $N$. What is interesting is the new condition for the determination of labor allocation across assets, $\alpha$. We will explore such distortions further in Section IV.
II. Implications for the Choice of Inputs: Capital and Labor

In this section we explore some basic implications of the previous section’s results to issues of input choice. It is useful to consider first the simple choice of labor when the firm has but one type of asset. After developing the economic intuitions for this setting, we proceed to examine the joint determination of labor and capital.

A. Labor

Consider the hiring decision of an at-will firm compared to the neoclassical firm. As already indicated in Result 4, we expect both overemployment and the resulting internal wage to equal \( w \). To illustrate this notion, consider the following example.

Example 1: Consider a firm with a linear demand curve for its final product, \( P(q) = a - bq \), and a linear production function \( q = n \). Thus, \( \pi(n) = (a - bn - w) n \). In particular, suppose that \( a = 25 \), \( b = 1 \), and \( w = 5 \): \( \pi(n) = (20 - n)n \). The neoclassical firm would hire \( n^* = 10 \) units of labor. The at-will firm’s payoff is given by

\[
\tilde{\pi}(n) = \frac{1}{n} \int_0^n (20 - s) ds = 10n - \frac{1}{2}n^2.
\]

The optimal choice of labor for this firm is \( n^* = 15 > n^* \). Here, as in other examples we present, when the firm has the freedom to employ additional labor, the equilibrium wage for the at-will firm equals the outside option of the employees in accord with Result 5. The solution is illustrated in Figure 1. As the example demonstrates, employment is higher by a ratio of 3:2 in the at-will firm. This result, it turns out, is independent of the underlying economic parameters in the neoclassical profit function whenever \( \pi \) is quadratic.9

\[
\text{PROOF:}
\]

Let \( \pi(n) = \phi_0 + \phi_1 n + \phi_2 n^2 \), with \( \phi_1 > 0 \), \( \phi_2 < 0 \). The neoclassical optimum is easily found to be \( n^* = -\frac{1}{2}(\phi_1/\phi_2) \). The at-will firm’s profits are

\[
\tilde{\pi}(n) = \frac{1}{n} \int_0^n \pi(s) ds = \phi_0 + \frac{1}{2}\phi_1 n + \frac{1}{3}\phi_2 n^2.
\]

Differentiating with respect to \( n \) and setting to zero yields \( \tilde{n}^* = -\frac{1}{2}(\phi_1/\phi_2) \). The resulting ratio is as stated.

The result that a firm may want to overemploy in order to increase its internal labor pool and thereby reduce the bargaining power of labor is quite intuitive.10

\[
\text{It should be noted that, although the ratio of employment levels is independent of the parameters of } \pi, \text{ it does generally depend upon the division of bargaining power. In our present case in which surplus is divided equally in pairwise negotiations, the factor is } \frac{2}{3}. \text{ More generally, the}
\]

9 It should be noted that, although the ratio of employment levels is independent of the parameters of \( \pi \), it does generally depend upon the division of bargaining power. In our present case in which surplus is divided equally in pairwise negotiations, the factor is \( \frac{2}{3} \). More generally, the

10 Other papers have discovered similar effects in which excess employment improves the firm’s ex post position with labor, but for very different reasons. In the share economy of Martin Weitzman (1983), for example, overhiring also serves to drive down wages, through a previous contracted (and binding) revenue-sharing agreement with employees. In contrast, in our setting the firm overhires prior to any wage agreement, cognizant of how this will affect future negotiations. Jonathan Feinstein and Jeremy Stein (1988) also consider a model in which the
B. Capital

Now consider the introduction of capital into the firm’s profit function. As indicated in Result 4, capital will be chosen to equate the average marginal profit of capital to zero, where the average is taken uniformly over all possible configurations of labor; that is, either \( \frac{1}{n+1} \sum_{i=0}^{n} \pi_{i}(i, k) = 0 \) or \( \frac{1}{n} \int_{0}^{n} \pi_{i}(s, k) \, ds = 0 \). Together with some information about the effect of capital on the marginal product of labor, we can determine whether there will be too little or too much capital vis-à-vis the neoclassical optimum. Formally, let \( k^{*}(n) \) and \( \hat{k}^{*}(n) \) represent the optimal levels of capital for the neoclassical and at-will firms, respectively, holding labor fixed at \( n \); let \( r \) denote the cost of capital. Then we have the following statement.

**THEOREM 2:** \( k^{*}(n) > \hat{k}^{*}(n) \) if \( F_{nk} > 0 \), and \( k^{*}(n) < \hat{k}^{*}(n) \) if \( F_{nk} < 0 \).

**PROOF:**

Fix \( n \). The neoclassical firm chooses \( k^{*}(n) \) such that \( F_{i}(n, k^{*}) = r \). The bargaining firm (in the discrete setting) chooses capital such that

\[
\frac{1}{n+1} \sum_{i=0}^{n} F_{i}(i, \hat{k}^{*}) = r.
\]

If \( F_{nk} > 0 \), then \( \frac{1}{n+1} \sum_{i=0}^{n} F_{i}(i, k) < F_{i}(n, k) \), and so \( \hat{k}^{*}(n) < k^{*}(n) \). Analogously, if \( F_{nk} < 0 \), then \( \hat{k}^{*}(n) > k^{*}(n) \). A proof of the continuous case proceeds along similar lines.

Intuitively, by increasing capital, the at-will firm obtains a weighted average of the marginal returns to capital of the various possible employment levels, ranging from 0 to \( n \) workers. When the marginal return to capital increases with labor, this weighted average is less than the marginal return with \( n \) employees (which is what the neoclassical firm obtains). Consequently, holding the number of employees \( n \) fixed, the at-will firm underemploys capital compared to the neoclassical firm. Likewise, when the marginal return to capital decreases with labor, the at-will firm hires more capital than the neoclassical firm for a

---

*Footnote:* The firm may overhire to reduce employee opportunism by reducing the employee’s outside option. In their setting, an employee who quits in the absence of employee redundancy takes a unique project opportunity with him. Employee redundancy, however, reduces the external option to this project insofar as the firm will be able to compete with the departing employee using the same information (which has been obtained by the redundant employee as well). This potential competition reduces the value of setting up a competing enterprise and therefore lowers the employee’s wage demands.
given \( n \). The following example illustrates the distortion.

**Example 2:** Let \( F(n, k) = 100\sqrt{nk} - n^2 - k^2 \), \( w = 1 \), and \( r = 1 \). The isoprofit contour plots of the neoclassical and at-will firms over choices of \( n \) and \( k \) are illustrated in Figure 2. As is clear, the contractual firm chooses a distorted capital–labor choice which underemploys capital and overemploys labor.

Although \( F_{nk} > 0 \) always implies that \( k^*(n) < k^*(\tilde{n}) \), this does not in turn imply \( \tilde{k}^*(\tilde{n}) < k^*(n) \). Unlike the above numerical example, it is possible that the excess labor hired by the at-will firm might raise the marginal product of capital so high as to induce \( \tilde{k}^* > k^* \). Nonetheless, for a common class of profit functions the capital–labor ratio is always distorted downward due to bargaining concerns.

**THEOREM 3:** Suppose that \( \pi(n, k) = n^\alpha k^\beta - wn - rk \), where \( \alpha + \beta < 1 \). Then

\[
\tilde{k}^* = \frac{1}{2} \left( \frac{k^*}{n^*} \right).
\]

**PROOF:**

Solving the first-order conditions for the neoclassical firm’s capital–labor ratio gives \( n^*/k^* = (\alpha/\beta)(r/w) \). The at-will firm’s payoffs are

\[
\tilde{\pi}(n, k) = \frac{1}{n} \int_0^n (s^\alpha k^\beta - w s) \, ds - rk
\]

\[
= \frac{n^\alpha k^\beta}{1 + \alpha} - rk - \frac{wn}{2}.
\]

The first-order conditions for \( \tilde{\pi} \) are

\[
\frac{\alpha}{n} n^\alpha k^\beta = \frac{w}{2} (1 + \alpha)
\]

\[
\frac{\beta}{k} n^\alpha k^\beta = r(1 + \alpha).
\]

Dividing yields \( \tilde{n}^*/\tilde{k}^* = 2(\alpha/\beta)(r/w) \).

This theorem illustrates a more general property of the at-will firm’s input choices. In the continuous labor setting, this wage-negotiable firm can be seen as acting as a neoclassical firm with a production function given by

\[
\tilde{F}(n, k) = \frac{1}{n} \int_0^n F(s, k) \, ds + \frac{1}{2} wn
\]

and maximizing \( \tilde{\pi}(n, k) = \tilde{F}(n, k) - wn - rk \). Note, however, the presence of \( w \) in the at-will firm’s induced production function \( \tilde{F} \) due to the wage-bargaining externality from additional labor. A higher external cost of labor \( w \) implies a greater return from using excess employment to hold down the internal wage demands. In Theorem 3 above, \( \tilde{F} = F/(1 + \alpha) + w n/2 \), which directly translates into a distorted capital–labor ratio of \( 1/2 \) due to the positive labor externality.

More generally, the possibility that an at-will firm may be mistaken for a neoclassical firm can lead to serious problems in the estimation of production and cost functions. As Theorem 3 above demonstrates, obtaining accurate estimates of the marginal rate of substitution between capital and labor requires knowledge of the bargaining environment (i.e., neoclassical or at-will). Moreover, because the market wage enters into the at-will firm’s induced production function, \( \tilde{F} \), standard duality-based estimation procedures such as estimating factor demand relations and using neoclassical optimization conditions to recover a firm’s production function are problematic. Furthermore, with an at-will firm, one must be careful about applying the standard properties of neoclassical cost functions. Although some properties (such as the monotonicity and concavity in factor prices) remain valid for the at-will firm’s cost function, other standard properties (such as first-degree homogeneity in factor prices) fail.

**C. Application: The Economic Incidence of Hiring Costs**

We now turn to a specific application of our methodology to the question of who actually pays the costs inherent in hiring (and perhaps training) a new employee. In the simplest bargaining model where the firm is free to vary
its employment level, we can think of our firm beginning negotiations with an exogenously given number of workers, say, \( n \). As negotiations progress, some workers may leave (or be sent away), resulting in a final number \( \tilde{n}^* \) who agree to the stable wage profile \( \tilde{w} \). The assumption that \( n \geq \tilde{n}^* \) begs the interesting and important questions of adjustment costs. In particular, it is reasonable to suppose that the firm may have to expend resources to train (or otherwise hire) the initial stock \( n \) well before wage negotiations are completed. In this subsection we first seek to establish who actually pays the training costs; this is a question of economic incidence. In the following subsection, we incorporate the model into an uncertain environment to explore the effect of stochastic shocks in economy-wide wages.

Suppose that there is some per-employee hiring cost of \( C \) which must be paid prior to production to have an employee available for wage negotiations and potential employment. The timing is straightforward. First, a sufficiently large number of potential employees, \( n \), simultaneously announce the fraction of the cost \( C \) (e.g., training costs) that they will bear in exchange for employment prior to negotiations. After observing these payments, the firm simultaneously chooses some of the employees to hire, \( n \), who then show up having undertaken the promised degree of "training." Unchosen employees go to work in the neoliberal sector for the wage of \( \tilde{w} \). At this point, we begin our familiar bargaining game resulting in an employment level of \( \tilde{n}^* \leq n \), an internal wage of \( \tilde{w}(\tilde{n}^*) \), and a profit of \( \tilde{\pi}(\tilde{n}^*) \).

It is immediate in the case of no uncertainty that the firm will never find it optimal to train more than \( \tilde{n}^* \) employees, as eventually all labor in excess of \( \tilde{n}^* \) will be dismissed. Hence, \( n = \tilde{n}^* \) in the deterministic setting. A more interesting question is how hiring costs \( C \) are divided in this setting. If employees offered to pay the entire hiring cost themselves, the firm would overemploy labor and drive the internal wage down to \( \tilde{w} \), leaving the employees with a net wage \( \tilde{w} - C \), which is less than their outside option, \( w \). On the other hand, if the firm paid for the hiring without payments by the employees, the firm would not hire as many employees, leaving an internal wage in excess of the outside option of the employees. In this case, employees would gladly incur a portion of the costs, \( C \), in order to get hired and earn rents. To find the equilibrium payment, we must find a fixed point \( (\tilde{n}^*, \theta) \), where \( \theta \) is the firm’s share of hiring costs, such that at this point the employees earn no excess rents in the firm given \( \tilde{n}^* \) and the firm optimally chooses \( \tilde{n}^* \), given hiring costs are fixed at \( \theta C \) per employee. Formally,

\[
(9) \quad \tilde{w}(\tilde{n}^*) - \tilde{w} = (1 - \theta)C
\]

\[
(10) \quad \tilde{n}^* \in \arg \max_n [\pi(n) - \theta C n].
\]

The first equation is the equilibrium condition for workers: their future excess rents exactly equal their contribution to the hiring cost, given the conjectured level of employment, \( \tilde{n}^* \). The second equation requires that the conjectured employment is indeed optimal from the firm’s viewpoint. For the continuous labor case, the first-order condition for (10) is \( \pi(\tilde{n}^*) - \tilde{\pi}(\tilde{n}^*) = \theta C \tilde{n}^* \). Since all output is realized in either profits or wages, \( \pi - \tilde{\pi} = n(\tilde{w} - \tilde{w}) \); substitution yields \( \tilde{w}(\tilde{n}^*) - \tilde{w} = \theta C \). Combining this expression with (9) produces our result.

THEOREM 4: In the deterministic wage-negotiation game with hiring costs, the firm and employee equally share in the payment of \( C \):

\[
\theta = \frac{1}{2}.
\]

Moreover, the firm’s optimal choice of \( \tilde{n}^* \) is equivalent to the choice of a firm which faces no hiring costs but an outside wage of \( \tilde{w} + C \):

\[
\tilde{n}^* = \arg \max_n \tilde{\pi}(n, \tilde{w} + C).
\]

Rather strikingly, this result indicates that the ex post split-the-difference bargaining over wages induces an equal ex ante split in who directly pays the hiring costs. The economic incidence of hiring costs is completely deter-
D. Application: Exogenous Shocks and Adjustment Dynamics

We now consider uncertainty over the future value of trained labor. If training is such that the firm cannot hire new employees in response to an exogenous shock until a period of production elapses, our setting generates a number of interesting adjustment dynamics. Specifically, if an exogenous shock raises the outside option, the firm can cut employment to the optimal ex post level. However, if a shock leads to a lower outside option, insofar as the firm cannot hire immediately, the firm’s ex post employment decision is constrained, and internal labor enjoys an increase in bargaining power. Thus, when there is uncertainty over future outside wages, there may be some option value to hiring additional workers, which must be balanced with training costs spent for workers who may not be utilized. In such a case, there will generally be some realized states in which the desired number of workers will exceed the number chosen up front. When hiring is thus constrained, the wage paid inside the firm will decrease by only a fraction as much as the outside shock, and workers will earn a wage premium.

We presently explore these issues in an example that builds on the hiring-cost model of the above subsection. We assume that training is valuable for one period of production only; hence, training only confers option value to the firm with regard to the uncertainty to be resolved in the current period. For this example, we take changes in the outside option \( w \) to be a reduced-form representation of the impact of shocks on the firm. Specifically, we assume that next period’s outside wage is distributed on \([w_0, w]\) according to a density \( g(w) \). Note that a positive (negative) shock to the firm’s marginal productivity would operate quite similarly to a negative (positive) shock to its outside option. It is important to emphasize that, in order to explore adjustments to such shocks, in this section we depart from our maintained assumption for comparative-statics results elsewhere, that the firm can hire new workers in response to changes in exogenous parameters. We define \( w^*(n) \) to be the critical value of \( w \) such that for all \( w < w^*(n) \), a firm with \( n \) workers would desire to hire more workers absent training costs, and for all \( w > w^*(n) \), the firm will prefer to dismiss some workers. Thus, \( w^*(n) \) is defined by the equation

\[
0 = n - \bar{w}^*(w^*(n)).
\]

When the realized outside wage is sufficiently high such that \( w \geq w^*(n) \), the firm will reduce its employment to the point where the internal wage equals \( w \). In such a state of the world, each worker (retained or dismissed) earns no excess rents. When \( w < w^*(n) \), on the other hand, all workers are retained and \( \bar{w}^*(n, w) \geq w \), with strictly excess rents realized in some states. Thus, the net gain from employment to a risk-neutral worker is the expected wage premium over the low-wage states. Equating this with the worker’s equilibrium contribution to hiring costs yields a direct analogue to equation (9) above:

\[
(11) \int_{w_0}^{w^*(n)} [\hat{w}(n, w) - w] dG(w) = (1 - \theta)C.
\]

Now consider the firm’s perspective. For a given stock of trained workers, \( n \), the firm’s profits can be decomposed into two regions of uncertainty. For \( w \leq w^*(n) \), the firm does not change its employment level from \( n \), and it has profits of \( \hat{\pi}(n, w) \). For \( w > w^*(n) \), the firm will optimally adjust its labor to \( \bar{w}^*(w) \), and profit is \( \hat{\pi}(\bar{w}^*(w), w) \). Taking expectations with respect to \( w \), we have the firm’s net profit given by

---

11 Using the framework in Stole and Zwiebel (1996), it is easy to generalize this result to show that the fractions of the ex ante training costs borne by the firm and the employees are always identical to the fractions of the surplus realized by the parties in the ex post pairwise wage negotiations, for any exogenously given bargaining split.

12 When \( n \) is so large that for all \( w, \bar{w}^*(w) < n \), we set \( w^*(n) = w_0 \); similarly, when for all \( w, \bar{w}^*(w) > n \), we set \( w^*(n) = w \).
The first-order condition for $n$ simplifies to

$$\int_{w_0}^{w_0} \frac{\partial \tilde{\pi}(n, w)}{\partial n} dG(w) = \theta C.$$ 

As in the previous section, $\partial \tilde{\pi} / \partial n = (1/n) \times (\pi - \tilde{\pi})$. Using the identity that $\pi - \tilde{\pi} \equiv n(w - \tilde{w})$, we can substitute wages for profits in the firm’s first-order condition. Combining the expression with (11), we have once again that the training costs are equally split.

**THEOREM 5**: In the $w$-uncertainty wage-negotiation game with hiring costs, the firm and employee equally share in the payment of $C$: $\theta = \frac{1}{2}$.

Intuitively, if training were costless, the firm would like to train the number of workers it would choose to have if the outside wage were $w_0$ in order to maintain the option of keeping the optimal number of workers for each state. However, the presence of hiring costs lowers the number of workers that the firm will hire up front, in much the way it did in the previous subsection without shocks. What is particularly striking is that the firm still covers exactly half of the training cost in equilibrium when uncertainty is present.

In addition, an at-will firm will exhibit interesting wage responses to external shocks. When a low-wage outcome is realized (i.e., $w < w^*$), the firm chooses to retain all trained workers, and the internal equilibrium wage exceeds the outside wage due to bargaining pressures. When instead a high-wage shock is realized, the firm scales back employment to the *ex post* optimum of a firm facing no hiring costs. In this case, the internal wage moves in step with the outside wage. This asymmetric movement results in wage premia arising in good times (i.e., low-$w$ or high-productivity states) and wages adhering to the outside wage during bad times.

### III. Applications of the Value of Front-Loading

In Section I we presented results indicating that the portion of the productive surplus that the firm receives as profits depends upon the global character of the neoclassical profit function, $\pi$, and not just its value (and margin) at the current level of employment. Specifically, we introduced the notion of a front-load factor to capture the degree to which the firm’s marginal products of labor are distributed “up front” for low numbers of workers. It is clear from Result 6 that, if two technologies are equally productive at $n$, the at-will firm will prefer, given $n$, that which is more front-loaded. Consequently, a firm may prefer a sub-optimal technology from a set of potential production functions if it is sufficiently front-loaded relative to the otherwise efficient choice. Thus, there will generally be distortions in the choice of technologies by an at-will firm, given that such a firm is prepared to sacrifice productive efficiency for strengthened bargaining power via higher front-loading.\(^{13}\) Additionally, although we have not formally modeled worker investments in human capital and other noncontractible employee actions, we could consider the preferences of workers for particular technologies using the front-load factor. Specifically, suppose that the individual workers have some ability to affect the front-load factor of the technology they work with, perhaps through investments in human capital that make them

\(^{13}\) The point that underlying concavity or convexity may directly affect the equilibrium payoffs in a bargaining game has been noted in several papers. Probably closest to our own result is that of Gilbert Skillman and Harl Ryder (1993) who demonstrate in a two-worker model that a firm may rationally choose a less efficient technology if the first margin is sufficiently high relative to the total return. In a related paper, Henrik Horn and Asher Wolinsky (1988) show that a firm’s preference for two smaller unions rather than one large union depends upon the substitutability of the labor groups in the given technology. Patrick Bolton and David Scharfstein (1993) argue that whether a firm will want to borrow from one creditor or two depends upon the returns to scale in the technology evaluated over two points. All of these papers insightfully note that the curvature of the underlying technology is key; by considering more general technologies, however, we are able to identify precisely the manner in which this curvature is important (i.e., front-loading).
more indispensable on the current margin, or perhaps by increasing their synergies with other workers at the cost of general productivity. In such a setting, the workers will generally distort their investments toward back-loading technology. Such investments have the effect of entrenching the workers and increasing their bargaining power. A manager, for example, may increase his negotiated wage by undertaking investments that raise his specific value to the firm and therefore back-load the firm’s profit function.\footnote{Andrei Shleifer and Robert Vishny (1989) and Aaron Edlin and Joseph Stiglitz (1995) consider these issues of “managerial entrenchment” in more specific settings. More generally, the literature on influence costs considers activity by agents in a firm that enhances their position to the detriment of the overall firm (see e.g., Paul Milgrom and John Roberts, 1990).}

The notion of front-load factors can also be applied to other interesting economic questions. Essentially, any variable that increases a firm’s front-load factor for any \( n \) will reduce the amount of overemployment the at-will firm chooses. We present three very different applications to illustrate the utility of this notion.

A. Application: The Importance of Firm-Specific Human Capital

An important issue related to our overemployment result involves asset specificity of human capital within the firm. To explore this topic, we suppose here that, in addition to the specific labor a firm can employ, it can also hire nonspecific labor in the external market at the wage of \( w \). We take the underlying neoclassical profit function to be given by \( \pi(n, m) \), where \( n \) is previously trained (firm-specific) labor within the organization and \( m \) represents the number of nonspecific employees. We assume that the at-will firm can go to the external labor market at any time during the negotiations and hire nonspecific labor at a cost of \( w \) per unit (and therefore only specific labor has holdup power). In this environment, the at-will firm’s profit is given by

\[
\hat{\pi}(n) = \frac{1}{n} \int_0^n \bar{\pi}(s) \, ds
\]

where \( \bar{\pi}(n) = \max_{m\geq 0} \pi(n, m) \). In the extreme case in which nonspecific labor is productively identical to specific labor, if \( n = n^* \) (specific labor is less than the neoclassical optimum), then \( \bar{\pi}(n) = \pi(n^*) \), and so \( \hat{\pi}(n) = \pi(n) \); there is no labor distortion. If, instead, nonspecific labor is useless (i.e., does not affect production) then trivially, \( \bar{\pi}(n) = \pi(n) \), and the distortion is as though there were no nonspecific labor. Generally, the distortion will be intermediate between the neoclassical setting and the pure wage-negotiation setting without an external labor supply. That is, the firm will still overhire relative to a neoclassical firm, but the extent of this overhiring will be lessened by the presence of nonspecific labor. The introduction of an external (albeit imperfect) labor supply at \( w \) has the effect of front-loading the neoclassical profit function. That is, even though \( \bar{\pi}(n) = \pi(n) \) for all \( n \geq n^* \) and therefore is equally efficient from a neoclassical viewpoint, \( \hat{\pi}(n) \) has a strictly higher front-load factor than \( \pi(n) \) and is strictly preferred by the at-will firm. An example illustrates the effect.

**Example 3:** Consider our previous setting in Example 1, but with an augmented production function, \( q = n + pm \), where \( p \in [0, 1] \) represents the extent to which outside labor is a substitute in the firm’s revenue product of labor. This gives us an expression for neoclassical profit equal to
\[ \pi(n, m) = (25 - n - \rho m)(n + \rho m) - 5(n + m). \]

Substituting for the optimal choice of \( m \) conditional on \( n \), we have

\[ \pi(n) = \begin{cases} 
(625 - 20n)\rho^2 + (20n - 250)\rho + 25 & \\
4\rho^2 & \text{for } n \leq 5(\rho - 1) \\
(20 - n)\rho & \text{otherwise.} \end{cases} \]

We consider both \( \rho = 0 \) and \( \rho = \frac{1}{2} \). When external labor substitutability goes from \( \rho = 0 \) to \( \rho = \frac{1}{2} \), the optimal choice of specific labor goes from 15 to 13.9, and the at-will firm’s profits increase (see Figure 3). The reduced asset specificity (and consequent greater front-loading) helps the firm and leads to fewer total employees.

The result that reduced asset specificity of labor improves the firm’s front-load factor, increases \( \pi \), and reduces the extent of overemployment is quite similar to the value of reversible capital investments. The underlying intuition for the above results is that if an employee quits (which does not occur in equilibrium), the firm can respond by going to the external labor market and hiring an imperfect substitute, which ultimately lessens the cost of the employee’s departure. Similarly, if the firm has the ability to change its capital stock (e.g., close down plants, switch to less labor-intensive production processes, etc.), it can minimize the loss from an employee quit. This front-loads the firm’s ultimate profit function, changing it from \( \pi(n, k) \) to \( \pi(n) = \max_k \pi(n, k) \). The analysis is almost identical to that presented above for the effect of labor-specificity. Notably, in the limit when capital is fully flexible and costlessly reversible, there will be no distortion in capital choice as labor can never hold up its value.

B. Application: Preferences for Unionization

We can think of employees’ (the firm’s) preferences over various unionization structures as a direct application of front-load factors: for two structures with identical output, the structure with the lower (higher) front-load factor is preferred. Here, we abstract from the many roles a union may serve and take it to mean solely a binding commitment by a fixed number of employees to bargain together for the same wage and to decide jointly whether to work as well.\(^{15}\) To this end, consider a firm with several employees in which employees can initially choose to deal with the firm either individually or as a collective group. When will the employees prefer fragmented negotiations to centralized ones? The answer to this question is immediate from Result 6.

A union has the effect of linearizing the labor-revenue-product function, which means that if the original technology was convex, a union makes it more concave (and hence more front-loaded since the endpoints are unchanged); likewise, if the original technology was concave, unionization makes it more convex. A union is desirable from the employees’ point of view (and hence undesirable from the firm’s point of view) whenever the underlying technology is concave, while the reverse holds true for convex technology.\(^{16}\) Insofar as employees can selectively choose when to commit to bargaining as a group and when instead to bargain individually, they can—as far as the bargaining outcome is concerned—effectively linearize concave regions of the production function while keeping convex regions fixed, thereby lowering the front-load factor and, for a fixed number of employees, raise their wages.

A simple numerical example with two employees illustrates the relationship between un-

---

\(^{15}\) It is important to note that this abstraction suppresses many of the important features of union–firm bargaining addressed in the labor literature (see e.g., Richard Freeman and James Medoff, 1984), which may conceivably interact with bargaining considerations we take up presently. For example, unions are often thought to be concerned with the number of members as well as wages.

\(^{16}\) A similar relationship between convexities and the desirability of unionization has been noted by Horn and Wolinsky (1988), although in a different bargaining environment with two employees (or two groups of employees).
F(n)

cient

suppose

APF(2)

simplicity,

xu

Figure

gical

only

employees

highly

union

thus

(and,

bargain

Note

that, although the notions of concavity and convexity provide the correct intuition in this application, in general the crucial condition is the notion of front-loading. One could easily imagine a particular organization of bargaining units such that it was neither concave nor convex compared to another structure. Nonetheless, a complete comparison always exists using front-load factors.

C. Application: The Effects of Competition on Observed Labor Efficiency

It is sometimes argued that monopolists are plagued with incentive problems, which are reduced by the presence of product-market competition.\textsuperscript{17} We provide an altogether different explanation, suggesting that such an empirical observation is consistent with the theory of an at-will firm where hiring decisions are second-best optimal. In short, the presence of competition fundamentally alters the curvature of the firm’s neoclassical profit function (and consequently its front-load factor), thereby affecting its bargaining relationship with employees and consequently its returns from overemploying labor. Depending upon the nature of competition, increased market pressure may increase or decrease the front-load factor of \( \pi(n) \), leading to either a decrease or an increase in employment. Below we consider two distinct forms of competition, illustrating each of these phenomena.

Because the at-will firm will always overemploy relative to the neoclassical firm, the relevant comparative static is the relative effect of competition on overemployment: \( \bar{n}*/n* \).

\textsuperscript{17} See, for example, Harvey Liebenstein’s (1966) discussion of X-inefficiencies. A similar story is frequently told about the cleansing effect of recessions: during a period of reduced demand, firms search for inefficiencies and waste within their organizations and dismiss employees with low marginal revenue products of labor.
We have already seen above in Theorem 1 that merely shifting or rotating a linear demand curve will have no effect on this ratio. When competition affects a segment of a firm’s customer base disproportionately, however, this can effect the relative proportion of employees that the two types of firms choose to hire. For example, a competitive fringe of firms may suddenly emerge which is prepared to sell unlimited quantities of output at their constant marginal cost, which is higher than the incumbent’s marginal cost. In this case, the incumbent’s effective (inverse) demand curve becomes horizontal for points at or above the fringe’s marginal cost, affecting the front-load factor of the underlying neoclassical profit function. Provided that the competitive fringe is sufficiently inefficient, potential competition will have no effect on the neoclassical firm (i.e., its monopoly price is already below the competitive fringe’s marginal cost). In this setting, we can unambiguously predict the effect that competition has on the ratio $n^*/n^*$.

Along these lines, we consider a firm with linear (inverse) demand, $P(q)$, and an outside wage $w < P(0)$. We assume that the production function is simply $q = n$, and therefore the monopolist’s profit is given by $\pi_m(n) = (P(n) - w)n$ when competition is not present; let $P^m$ denote the corresponding optimal monopoly price. When the competitive fringe is introduced, it is willing to supply unlimited quantities at the price of $\bar{p} < P(0)$. We assume that $\bar{p} > P^m$ so that the neoclassical firm’s hiring decision is unaffected by competition. Let $\pi_c(n)$ represent the neoclassical profit function of the firm after competition is introduced. A numerical example with linear demand nicely illustrates the effect of a competitive fringe on these functions, and hence on the at-will profits $\pi_m(n)$ and $\pi_c(n)$.

**Example 4:** Let $P(q) = 1 - q$, $w = 0$, and $\bar{p} = 0.6$. Before and after the introduction of competition, the neoclassical firm chooses $n^* = 0.5$. The at-will firm’s choice differs, however: before, $\bar{n}_u = 0.75$; while after, it increases to $\bar{n}_u = 0.78$.

The presence of competition lowers some of the previously high inframargins by restricting the firm’s ability to raise prices in the event of a forced output reduction. Specifically, the introduction of the competitive fringe effectively linearizes the neoclassical production function over the domain $[0, P^{-1}(\bar{p})]$, as indicated in Figure 5. From Section I, we know that such a transformation of the neoclassical profit function, which lowers its value over some interval without affecting its endpoints, reduces the front-load factor at the endpoint and induces a higher employment level from the at-will firm. Thus, $\pi_m$ is more front-loaded than $\pi_c$ for all $n \geq P^{-1}(\bar{p}) = 0.4$. Accordingly, a firm facing competitive pressures will increase its employment rolls in order to reduce the internally bargained wage. The neoclassical firm, however, does not change its hiring decision, as competition only affects the inframarginal profits. Intuitively, competition raises labor’s bargaining strength in an at-will firm because under competition a reduction in output can push the firm into an employment region that is even less profitable than before. Previously, the firm could respond to low output with a higher price, mitigating the output loss; with competition, such mitigation is constrained. This intuition generalizes as follows.

**THEOREM 6:** Suppose that $\bar{p} > P^m$. Under uniform pricing, the introduction of a competitive fringe at price $\bar{p}$ increases the employment level of the incumbent at-will firm compared to the neoclassical firm (both absolutely and proportionally).

**PROOF:**

Let $n^*$ be the neoclassical optimum. Let $\bar{n}$ be the level of employment that generates
a price of \( \bar{p} \): \( \bar{n} = P^{-1}(\bar{p}) \). By assumption \( \pi_m(n) = \pi_c(n) \) for all \( n \geq \bar{n} \), and therefore the neoclassical firm chooses the same \( n^* \) regardless of whether there is a competitive fringe. Competition, however, linearizes \( \pi_c(n) \) over the domain from 0 to \( \bar{n} \), connecting 0 with \( \pi(\bar{n}) \). Because the integral over \([0, \bar{n}^*_m]\) of \( \pi_c(n) \) exceeds that for \( \pi_c(n) \), the front-load factor at \( \bar{n}^*_m \) is lowered below unity when competition is introduced. Hence, the conditions for an at-will firm’s optimum require that \( \bar{n}^*_m < \bar{n}^*_c \).

We now turn to a related form of competition in which the incumbent firm perfectly price-discriminates (i.e., sets a different price for each unit sold) and competition, as before, consists of a competitive fringe willing to supply unlimited quantity at the fringe’s higher marginal cost, \( \bar{p} \). In such a setting, we will see the above result is reversed; here the introduction of a competitive fringe decreases the relative employment of the at-will firm. Consider first a simple numerical example.

**Example 5:** Let \( P(q) = 1 - q \), \( w = 0 \), and \( \bar{p} = 0.25 \). Before and after the introduction of competition, the neoclassical firm chooses \( n^* = 1 \). The at-will firm’s choice differs, however: before, \( \bar{n}^*_m = 1.5 \); while after, \( \bar{n}^*_c = 1.39 \).

Intuitively, the effect of introducing a competitive fringe with higher marginal costs on the neoclassical production function can be decomposed into two components. First, the neoclassical profit function shifts downward by a fixed constant equal to the lost rents previously realized from price-discrimination on all high-demand consumers who were willing to pay in excess of \( \bar{p} \). Such a constant downward shift in profits, represented by the dashed extension to the \( \pi_c(n) \) curve in Figure 6, does not affect the production decisions of either the neoclassical or the at-will firm; it is equivalent to an increased fixed cost of production. However, this shift overrepresents the effect of competition on the neoclassical firm’s profits over the interval \([0, P^{-1}(\bar{p})] \), since at these low levels of production not all this profit is being realized in the first place, as only a fraction of the high-demand consumers are being serviced. Indeed, up to the point \( P^{-1}(\bar{p}) \), the neoclassical firm’s profits are once again linearized by competition; the firm obtains a fixed price \( \bar{p} \) from all customers over this range. Hence, the actual neoclassical profit function with the competitive fringe is given by the solid representation of \( \pi_c \) in Figure 6. This second shift, however, is the reverse of that induced by the introduction of a competitive fringe in the uniform-pricing case; it consists of a shift that raises the value of the profit function over the range \([0, P^{-1}(\bar{p})] \) without affecting its endpoint. This shift consequently increases the front-load factor at the endpoint and induces a lower employment level from the at-will firm, while leaving the neoclassical firm’s production level (which we have assumed exceeds \( \bar{p} \)) unchanged. In contrast to the previous case, the introduction of a competitive fringe under price discrimination lowers the neoclassical profit function more at high levels of production than low levels. With price discrimination, the margins that are eroded by such a competitive fringe were more susceptible to employee holdup than without price discrimination, and consequently, the introduction of this fringe (and the subsequent loss of these margins) makes excess labor relatively less valuable to the firm. Our result once again generalizes.

**THEOREM 7:** Assume that firms perfectly price-discriminate. The introduction of a competitive fringe decreases the employment level of the incumbent at-will firm compared to the neoclassical firm (both absolutely and proportionally).
PROOF:

Let \( \hat{n}^*_{m} \) be the original at-will monopolist’s optimal employment level. Since Results 4 and 6 together indicate that front-load factors are driven up to 1 at the optimum, it follows that \( \int_{0}^{\hat{n}^*_{m}} s \pi'_{m}(s) \, ds = 0 \). Over the region of employment where \( \hat{\rho} \) does not bind, \( \pi'_{m}(s) = \pi''_{m}(s) \); the profit functions differ only by a fixed constant. Over the region where the outside competitive price of \( \hat{\rho} \) constrains the firm’s offered price, \( \pi'(s) < \pi''_{m}(s) \). Thus, \( \int_{0}^{\hat{n}^*_{m}} s \pi'(s) \, ds < 0 \). This implies that the front-load factor under competition at \( \hat{n}^*_{m} \) exceeds unity, implying \( \hat{n}^*_{m} < \hat{n}^*_{m} \). Furthermore, since the neoclassical firm’s output decision is unaffected by potential competition, the employment in the at-will firm decreases relative to the neoclassical firm.

The introduction of price competition reduces the holdup power of labor, thereby raising the front-load factor. The result is that the firm overemploys to a lesser extent. It is worth remarking that the above effect of perfect price-discrimination continues to hold when discrimination becomes less than perfect. For example, consider the setting in which a firm practices third-degree price discrimination and sells to two markets, one high demand and one low demand. The competitive fringe directly enters the high-demand market, ignoring the less profitable low-demand consumers. Here a similar effect emerges from the lowering of the high-demand market margins, and the firm will desire fewer workers. Of course, there is also a countervailing force similar to the uniform-price setting above, leading in the opposite direction. Which effect dominates depends upon the underlying market demand curves. In general, competition can either enhance or reduce the bargaining position of labor and therefore can either induce an increase or decrease in the overemployment magnitudes, depending upon inframarginal effects. The two important effects are nonetheless clear: competition can reduce front-loading and therefore increase employment when competitive pressures restrict the firm’s ability to respond to labor holdout by raising prices; competition can alternatively increase front-loading and reduce employment when prior to competition the firm is not able to adjust prices further to mitigate the consequences of labor holdout (because the firm is already fully extracting consumer surplus).

IV. Organizational Design

We now turn to applications where the firm’s production function will have multiple labor arguments, representing heterogeneous labor.\(^8\) While we will generally speak of this differentiated labor in terms of employees assigned to different assets, the generality of these results should be emphasized. In particular, we could just as soon be talking about \( \text{ex ante} \) differentiated employees or employees assigned to different “teams.” What will be important, however, is the presumption that, upon bargaining, these employees cannot be reassigned across “assets”; that is, due to training or an \( \text{ex ante} \) differentiation, the “assets” that employees are assigned to are fixed. In Subsection IV-E we consider cross-training and examine the consequences of allowing firms to vary this assignment. Reviewing notation, we will consider a firm with production function \( F(n_{A}, n_{B}) \), where \( n_{A} \) and \( n_{B} \) are the amount of labor the firm has to work on assets \( A \) and \( B \) respectively. We let \( N = n_{A} + n_{B} \) represent the total scale of production, and let \( \alpha = (\alpha_{1}, \alpha_{2}) \) represent the labor-allocation decision, where \( \alpha_{i} = n_{i}/N \in [0, 1] \).

A. Economies of Scope

We begin by demonstrating that employee bargaining induces simple scope distortions across different groups of labor that are similar to the scale distortions explored in Section III. This can be seen most clearly in a simple example holding labor decisions \( (N \text{ and } \alpha) \) fixed. Let \( n_{A} = 1 \) and \( n_{B} = 1 \), and \( w = 0 \) for both employees. This setting could, for example, depict two unions that represent minimal bargaining units. We define the underlying scope economy naturally by \( \Delta sF(1, 1) \equiv F(1, 1) - F(0, 0) - F(0, 1) + F(0, 0) \). \( \Delta sF \) indicates the total additional revenue product of labor produced from the synergies of the two pro-

\(^8\) In all of our examples two arguments will suffice, though it is straightforward to generalize.
productive groups. A neoclassical wage-taking firm would choose technology, \( F \), so as to maximize \( \pi = F(1, 1) \). Result 3, however, indicates that

\[
\bar{\pi} = \frac{1}{3} \left[ F(0, 0) + \frac{F(1, 0)}{2} + \frac{F(0, 1)}{2} + F(1, 1) \right]
\]

\[
= \frac{1}{2} \left[ F(0, 0) + F(1, 1) \right] - \frac{1}{6} \Delta F.
\]

Note that, instead of simply choosing technology to maximize total output \( F(1, 1) \), the at-will firm cares about \( F(0, 0) \) and \( \Delta F \) as well. The \( F(0, 0) \) is just our familiar front-load effect; technologies which produce more from earlier margins are favored over equally efficient alternatives. The \( \Delta F \) term, however, indicates that holding overall output fixed, the at-will firm will also disfavor economies of scope. Intuitively, both the employee on asset \( A \) and the employee on asset \( B \) are essential to realize the economy of scope, and consequently both extract surplus from this in bargaining. If, instead, economies of scope were 0, only one employee would be essential for any particular margin. This result should not be confused with the incorrect statement that at-will firms do not like economies of scope: they do, insofar as they contribute to total output (just as with economies of scale). However, given a certain total output, such firms would prefer this output to be generated with economies of scope (and scale) that are as small as possible.

One potential application of this result is to Williamson's (1975) puzzle of selective intervention, which asks why mergers are not always profit-increasing given that all previous independent activities can be duplicated and the combined entity can selectively intervene to capitalize on any synergies. Such logic would imply that any pair of firms that impose externalities on one another could benefit by merging. Consider, however, the merger of two firms that impose positive externalities upon one another when bargaining concerns are present. In our setting, the employees can now make claims on the positive externalities that previously existed and were unclaimable. Such considerations of intrafirm bargaining pressures may serve as an important countervailing force, explaining why firms that impose externalities upon one another may choose to remain separate entities.

B. Input Misallocation Across Divisions

We now consider a simple example concerning the labor-allocation choice \( \alpha \) of an at-will firm with a given technology, \( F \), and number of employees, \( N \). We demonstrate here such a firm’s incentive to distort its \textit{ex ante} choice of employee assignment to assets or groups (or equivalently, what types of employees to hire), given the bargaining that will later ensue.

Let \( N = 2, w = 0 \) for all employees, and consider a technology, \( F \), such that \( F(0, 0) = 0, F(1, 0) = 0, F(0, 1) = 3, F(1, 1) = 4, F(0, 2) = 3, \) and \( F(2, 0) = 5 \). With one employee, both the neoclassical and the at-will firm would allocate the employee to asset \( B \). With two employees, the neoclassical firm would clearly allocate both workers to asset \( A \) yielding a profit of \( \pi(2, 0) = 5 \). In particular, its preferences over the three possible configurations of two employees is given by \( \pi(2, 0) > \pi(1, 1) > \pi(0, 2) \). On the contrary, however, the at-will firm would place both employees with asset \( B \). Specifically, \( \bar{\pi}(2, 0) = \frac{5}{3} + \frac{1}{4}(0) + \frac{1}{4}(5) = \frac{8}{3}, \bar{\pi}(1, 1) = \frac{3}{2}(3) + \frac{1}{4}(0) + 0 + 4] = \frac{11}{6} \), and \( \bar{\pi}(0, 2) = \frac{5}{3} + \frac{1}{4}(3) + \frac{1}{4}(3) = \frac{8}{3} \). In these three labor configurations, the wages of the employed workers are \( \bar{w}_A(2, 0) = \frac{5}{3}; \bar{w}_A(1, 1) = \frac{1}{4} \) and \( \bar{w}_B(1, 1) = \frac{11}{6} \); and \( \bar{w}_B(0, 2) = \frac{5}{2} \). It is notable that the order of preferences for the at-will firm, \( \bar{\pi}(2, 0) < \bar{\pi}(1, 1) < \bar{\pi}(0, 2) \), is exactly the opposite of that for the neoclassical firm.

C. Hierarchical Design

We now turn to a continuous labor application, dealing with hierarchies, in which we

\[19\] Here, we implicitly assume that an employee’s ability to hold up his own firm is greater than that of one with which his firm interacts. This could conceivably be justified by recourse to transaction costs.
demonstrate how hiring distortions across differentiated labor can arise in a natural setting. Suppose firms hire both high-level managers and lower-level workers who bargain with the owners of a firm. We let \( m \) and \( n \) represent the number of managers and workers hired, and \( w_m \) and \( w_n \) represent their respective outside wages, where we assume \( w_m > w_n \). We also assume that production takes the following form:

\[
F(n, m) = H\left( n \phi \left( \frac{m}{n} \right) \right)
\]

with \( \phi' > 0, \phi'' < 0, H' > 0, \text{ and } H'' < 0; \phi \) can be interpreted as the efficiency of the workers, which depends on the ratio of managers to workers. Managers are only beneficial insofar as they increase workers’ efficiency, and they do this at a decreasing marginal rate. From here, it follows that if the firm hires \( N = n + m \) total employees at a ratio of \( \alpha_m \) managers to total employees,

\[
\pi(\alpha_m, N) = H\left( (1 - \alpha_m)N \phi \left( \frac{\alpha_m}{1 - \alpha_m} \right) \right) - w_mN\alpha_m - w_nN(1 - \alpha_m).
\]

Letting \( \alpha_m^*(N) \) represent the neoclassical optimal allocation decision for a given \( N \), it follows (by definition) that

\[
\pi_{\alpha_m}(\alpha_m^*(N), N) = 0 \quad \forall N.
\]

As a preliminary step to our result, we seek to understand how \( \alpha_m^*(N) \) varies with total scale, \( N \), in this hierarchical firm. Assuming an interior solution and a quasi-concave neoclassical profit function, we can totally differentiate the above first-order condition with respect to \( \alpha_m \) and \( N \) and solve for \( da_m^*(N)/dN \).\(^{20}\) After some substitution with the first-order condition, we have

\[
\frac{da_m^*(N)}{dN} = - \frac{H''(\cdot)(1 - \alpha_m)(w_m - w_n)\phi}{H'(\cdot)\pi_{\alpha_m\alpha_m}} < 0.
\]

Intuitively, the dependence of the neoclassical optimal allocation \( \alpha_m^* \) on \( N \) follows from the lack of homotheticity in the underlying profit function. This lack of homotheticity in turn implies that \( \tilde{\alpha}_m^*(N) \neq \alpha_m(N) \).\(^{21}\) In our hierarchy setting, because \( da_m^*(N)/dN < 0 \), there will be an unambiguous distortion of the at-will firm toward higher levels in the hierarchy. Thus, top-heavy hierarchies emerge.

**THEOREM 8:** Suppose in a two-asset setting that \( \tilde{\pi} \) is strictly concave in \( m \) and \( n \), and \( da_m^*(N)/dN < (>) 0 \ \forall N \). Then holding \( N \) fixed, the resulting choice of the at-will firm will distort its employment decisions toward (away from) asset \( m \): \( \tilde{\alpha}_m^*(N) > (<) \alpha_m^*(N) \).

**PROOF:**

If \( da_m^*(N)/dN < 0 \ \forall N \), then \( \pi_m(\alpha_m N, (1 - \alpha_m)N) / \pi_n(\alpha_m N, (1 - \alpha_m)N) \) is decreasing in \( N \). Integrating, together with the first-order condition for \( \alpha \) for the neoclassical firm, implies that

\[
\int_0^N \pi_m(\alpha^*(N)s, (1 - \alpha^*(N))s) \, ds > \int_0^N \pi_n(\alpha^*(N)s, (1 - \alpha^*(N))s) \, ds.
\]

From equation (8) and the concavity of \( \tilde{\pi} \), this implies that \( \tilde{\alpha}_m^*(N) > \alpha_m^*(N) \).

Another more straightforward method of obtaining the result that an at-will firm may have top-heavy hierarchies follows from a direct consideration of bargaining power. Sup-

\(^{20}\) Alternatively, it is straightforward to show that \( \pi(\alpha_m, -N) \) exhibits the single-crossing property in \( (\alpha_m, N) \), from which the sign of \( da_m^*(N)/dN \) follows immediately (see Paul Milgrom and Christine Shannon, 1994).

\(^{21}\) In Stole and Zwiebel (1996), we examine a weaker notion of \( \alpha \)-homotheticity which is equivalent to having a linear expansion path along the given vector \( \alpha \); if a function is \( \alpha \)-homothetic for all vectors \( \alpha \), it is homothetic. This weaker notion of homotheticity is the necessary condition for \( \tilde{\alpha}_m^*(N) = \alpha^*(N) \) for all \( N \) (i.e., the allocation of the at-will firm is the same as the neoclassical firm), which can be seen immediately from equation (8).
pose that the higher tier splits the difference in pairwise bargaining sessions (as before), but the lower tier has no such power. Given this assumption regarding bargaining power, the at-will firm will continue to overemploy its senior managers (as always), but there is no reason to overemploy the second tier. Indeed, the second tier will be chosen at the standard neoclassical profit maximizing level, conditional on the overemployment of high-level managers. As above, we have immediately that $\bar{\alpha}_{m}(N) > \alpha_{m}(N)$.

D. Capital Budgeting

We now show that a simple reinterpretation of variables provides an interesting application to capital budgeting. We consider the problem of a firm choosing how much capital $k_A$ and $k_B$ to allocate to projects $A$ and $B$ at the cost of $r$ per unit of capital.

First consider an extremely basic example with discrete labor allocation. In particular, suppose that each project has a single essential employee (or equivalently, consists of employees who can commit to bargaining together as a unit), with outside wage $w = 0$. In addition, we presume that, if the employee for an asset does not reach an agreement with the firm and departs, capital from this asset is useless; that is, $F(0, 1, k_A, k_B) = F(0, 1, 0, k_B)$ and $F(1, 0, k_A, k_B) = F(1, 0, k_A, 0)$, where the first two arguments of $F$ are the numbers of employees on the assets (taken to be 0 or 1) and the last two arguments are the capital inputs to each asset. With these assumptions, we can redefine production to depend only on capital and whether the employee for the asset is working; that is, $G(k_A n_A, k_B n_B) = F(n_A, n_B, k_A, k_B)$, where $n_A, n_B \in \{0, 1\}$.

Given employee bargaining, provided that the firm operates both projects with some capital (i.e., there is an interior solution), the firm must solve the following capital-budgeting problem:

$$\max \bar{\pi} = \frac{1}{2} G(0, 0) + \frac{1}{6} G(0, k_B) + \frac{1}{6} G(k_A, 0)$$
$$+ \frac{1}{3} G(k_A, k_B) - r(k_A + k_B).$$

The distortion relative to the neoclassical firm, which of course only cares about $G(k_A, k_B) - r(k_A + k_B)$, is similar to that discussed above with economies of scope. In particular, the at-will firm will have a bias toward investing in the project that provides higher returns when the other project is not undertaken. This will lead to smaller scope economies for the employees to hold up in bargaining. Also, it should be clear that, holding total output fixed, the at-will firm would prefer a technology $G$ that yields fewer scope economies between the capital expenditures on each project.

Now consider a firm that is initially funding project $A$ and considering project $B$ as well. We will take capital levels as fixed (i.e., a certain fixed amount of capital is necessary for the project) and ask only what the firm’s incentive is to undertake a new project, $B$. To a neoclassical firm, project $B$ offers a return of $R_B = G(k_A, k_B) - G(k_A, 0)$ at a cost of $rk_B$. To an at-will firm, the net return from the project is

$$\frac{1}{3} R_B + \frac{1}{6} [G(0, k_B) - G(0, 0)] - rk_B.$$

Insofar as production under the two projects is complementary (i.e., $G_{12} > 0$), it follows that $R_B > G(0, k_B) - G(0, 0)$, and consequently the at-will firm will be less likely to undertake the new project. If the two projects are in-production substitutes (i.e., $G_{12} < 0$), the at-will firm obtains an additional strategic benefit from undertaking the new project which may be sufficiently great so as to induce an at-will firm to adopt the project even though a neoclassical firm would not. Mathematically, this will occur whenever $G(0, k_B) - G(0, 0)$ sufficiently exceeds $R_B$ such that

$$\frac{1}{3} R_B + \frac{1}{6} [G(0, k_B) - G(0, 0)] > rk_B > R_B.$$
We now turn to the continuous case, in which the firm can arbitrarily allocate labor and capital to its two projects. As before, we let \( F(n_A, n_B, k_A, k_B) \) represent the underlying technology. This implies the first-order conditions:

\[
\frac{1}{N} \int_0^N (F_{n_A}(\tilde{\alpha}^*s, (1 - \tilde{\alpha}^*)s, k_A, k_B) \quad - F_{n_B}(\tilde{\alpha}^*s, (1 - \tilde{\alpha}^*)s, k_A, k_B)) s ds = 0
\]

\[
\frac{1}{N} \int_0^N (F_{k_A}(\tilde{\alpha}^*s, (1 - \tilde{\alpha}^*)s, k_A, k_B)) ds = r
\]

\[
\frac{1}{N} \int_0^N (F_{k_B}(\tilde{\alpha}^*s, (1 - \tilde{\alpha}^*)s, k_A, k_B)) ds = r.
\]

The latter two equations suggest that the less sensitive \( F_k \) is to changes in \( N \), the less the allocational distortion. If \( F_{n,k_A} \approx 0 \) while \( F_{n,k_B} \approx 0 \), there will be an allocational bias in favor of project \( A \), which is less sensitive to labor holdups.

As we discussed earlier in Section III, when capital is reversible, the holdup problem for the benefits of capital is minimized. Therefore, if a firm has two projects and one of them has more reversible investment (higher resale value), this project will obtain higher relative funding than in the neoclassical case. Thus, despite the lack of any uncertainty in our setting, the option value of reversible capital is beneficial and attracts greater internal capital on the margin. The option value derives from the threat it confers in the bargaining game. While in equilibrium this option will never be exercised, its potential off-the-equilibrium-path usage can lead to important effects on negotiated profits.\(^{23}\)

E. Benefits of Cross-Trained Employees

Up to this point, we have presumed that, when faced with a multiasset or multigroup technology, the firm must first assign employees to their respective groups, and then bargain, taking this assignment as fixed. This indeed is likely to be reasonable in many settings under the training-cost interpretation of employee holdup power; the firm hires employees to work on different assets and trains them appropriately prior to production (and wage determination). However, such an interpretation begs the question of what are the potential benefits to the firm of cross-training employees, so that the firm may freely reassign them while bargaining. Our setting seems well suited to consider this and related questions. Indeed, the notion that firms cross-train employees for the purpose of replacing holdouts (or at least to have this threat during negotiations) is consistent with widespread anecdotal evidence. Nonetheless, formal modeling of this set of questions has not, to our knowledge, received much attention in the academic literature.

In order to address the potential bargaining benefits of cross-training, we must first be a bit more careful in specifying timing in our bargaining game. In particular, we will presume that, after any bargaining arrangement is agreed upon, the firm can reassign employees across assets. In effect, during the time when a firm is negotiating wages, we do not allow it to commit to not moving employees across assets for which they are qualified. Such redepolyments, of course, will be anticipated by employees and will affect the nature of the bargaining. Thus, a stable outcome with \( N \) employees will now consist of a wage, profit, and employee assignment profile \( \{ \tilde{w}(i), \tilde{\pi}(i), \tilde{\alpha}(i) \}_{i \in N} \), such that no one can benefit from renegotiations and, additionally, the firm cannot benefit from reassigning employees across assets for any \( i \).

First consider the case in which all \( N \) employees are cross-trained. For simplicity, we consider continuous labor. The only stable outcome will have \( \tilde{\alpha}(i) = \alpha^*(i) \forall i \leq N \), the neoclassical optimum. That is, if a bargaining agreement with \( i \) employees is obtained, everyone recognizes that, prior to production, the

\(^{23}\) Our capital-budgeting considerations are similar to those in Shleifer and Vishny (1989), where distortions in internal capital allocation arise due to incentives of individual managers to undertake investments that increase their own marginal value to the firm. Here, instead, the firm effectively distorts capital allocations in choosing investments that lessen managers’ holdup power through front-loading.
firms will choose to assign its $i$ employees in a manner that maximizes output $F$ (i.e., precisely the neoclassical optimal manner). Note that this is quite similar to the capital decision of the at-will firm when capital is reversible; all recognize that the choice will be that which maximizes output $F$. Thus, the profit for the at-will firm no longer is given by the neoclassical profits averaged along the ray from the origin to the allocation decision. Instead, if negotiations break down with some employees, the firm reallocates employees in the manner that maximizes $F$. Consequently, profits are instead given by averaging neoclassical profits along this neoclassical optimal labor expansion path. Specifically,

$$
\tilde{\pi}_{\text{CT}}(N) = \frac{1}{N} \int_0^N \{ F(s \sigma^*(s)) - sw_{\text{CT}} \} \, ds,
$$

where $\tilde{\pi}_{\text{CT}}$ and $w_{\text{CT}}$ are profits for the firm with cross-trained employees and the outside option for these employees, respectively. As usual, such a firm will hire employees until wages are driven down to the outside option, but here we must be careful in specifying what the outside option is for a cross-trained employee. Initially suppose that cross-training does not alter an employee’s outside option; that is, $w_{\text{CT}} = w$. Then, comparing equation (13) to (6) and noting that for all $s$, $F(s \sigma^*(s)) \equiv F(s \bar{\sigma}*)$, and also that the cross-trained firm could choose the labor choice of the non-cross-trained firm $\bar{N}*$, it follows that, $\tilde{\pi}_{\text{CT}} \equiv \tilde{\pi}$. A simple geometrical interpretation provides the intuition. While the at-will firm without cross-trained employees can only choose the endpoint of the (linear) ray in neoclassical profit space along which to average profits, the cross-trained firm can choose the entire labor-expansion path over which to average neoclassical profits. Thus, the flexibility that cross-training provides under bargaining translates mathematically into the ability to choose the optimal labor-expansion path over which to integrate neoclassical profits.

Of course, in many circumstances cross-training does not come for free for the firm. The cross-trained worker is likely to be able to realize at least the highest reservation wage associated with her various component skills, and perhaps a higher wage if other firms also demand employees with similar cross-training. Additionally, the training itself may be costly to the firm. Consequently, the firm may choose to cross-train only a fraction of its workforce, and may only do so across a subset of its required tasks.

One can solve for the optimal level of cross-training in such settings by expanding the dimension of our production function to allow for cross-training. For ease of exposition, we demonstrate this only for the two-asset case here; generalizing the technique to any finite number of assets is straightforward (though solving the program can potentially be quite cumbersome). In particular, consider our two-asset firm with production function $F(n_A, n_B)$, with the ability to cross-train employees at the additional cost of $t$ per employee, which in turn raises such an employee’s reservation wage to $w_{\text{CT}}$. It is readily apparent that this problem is identical to our standard multiasset problem, if we add a third dimension for cross-trained employees to our production function, in the process taking into account added training costs. In particular, we define

$$
\bar{F}(n_A, n_B, n_{\text{CT}})
= \max_{0 = p = n_{\text{CT}}} F(n_A + p, n_B + n_{\text{CT}} - p) - n_{\text{CT}}t
$$

$$
w_{\text{CT}} = (w_A, w_B, w_{\text{CT}}).
$$

Note that, in this manner, cross-training is essentially treated as a 'new asset,' which happens to have a very particular interaction with other assets in the production function. Treating cross-training as such, it then follows that the firm’s optimal labor and cross-training choice is equivalent to that of a multiasset firm with this additional asset in production. Thus, the optimal labor, allocation and cross-training decision follows from solving,

$$
\max_{a, N} \tilde{\pi}(n)
= \frac{1}{N} \int_0^N \{ \bar{F}(s \sigma) - s \sigma \cdot w_{\text{CT}} \} \, ds.
$$
Note that by augmenting $F$ with a dimension for cross-trained employees, we have transformed the problem from one in which any labor expansion path could be selected (as in the case when all employees are cross-trained) back into choosing the optimal ray along which to average neoclassical profits. In doing so, however, we have to add an extra dimension along which to choose the optimal ray.\textsuperscript{24} This technique converts the rather complicated problem of choosing the optimal degree of cross-training in our model to a standard multiasset problem, which we have already solved.

V. Extensions and Concluding Remarks

In this paper we consider the effect of intrafirm bargaining between employees and the firm on a wide range of organizational questions. Taken as a totality, these applications demonstrate the broad scope of interesting and relevant questions concerning firm behavior on which a careful examination of wage determination has bearing. These questions include not only the immediately apparent topics of hiring and input and technology choice, but also issues related to training and cross-training, firm-specific capital, unionization, the effect of competition, hierarchical design, and capital budgeting.

As such, we see the simple bargaining model employed here (and developed with more generality and rigor in Stole and Zwiebel [1995]) as a framework on which to build a general theory of organizations. While the applications presented in this paper are numerous and diverse, we believe that there are many additional organizational questions for which intrafirm bargaining is likely to be important. For brevity, we confine ourself to discussing just one here, albeit, one we find particularly important and interesting.

Specifically, while we have considered bargaining between a “firm” and its employees under a given determination for bargaining power, it seems promising to endogenize what is meant by “the firm” and the determination of bargaining power. Consider in particular a conception of the firm more along the lines of Jensen and Meckling’s “nexus of contracts,” where the firm is taken to consist of its employees and the capital they possess together with explicit and implicit contractual arrangements between them. Our paper in turn suggests the questions of how such employees interact in the presence of nonbinding labor contracts, and the implications of such interactions on the division of surplus.

A natural manner to approach these questions, which we are exploring in current research, is to modify our setting to “endogenize the firm.” In particular, consider allocating the role of the “firm” in our setting to the employees. That is, rather than exogenously choosing the individual who makes input decisions, consider the endogenous allocation of this power to employees. Firm “profits” would be divided among different employees (or perhaps groups of employees) on the basis of their relative importance in a bargaining game within the firm. The bargaining position of a group of employees would in turn likely depend on the production of different groups.

Embedding such a setting in a dynamic framework is likely to yield interesting implications for the relationship between different groups in an organization. Envisioning the organization as an evolving political contest between differing constituencies, who benefit from joint production but have conflicting goals on the division of surplus, seems to correspond well with intuitive notions of many organizations. Consider for example an academic department or governmental bureaucracy. Within such organizations, different groups are likely to complement one another in production, while at the same time to differ

\textsuperscript{24} When there are $M > 2$ assets in the production function, the same technique can be employed to obtain the optimal cross-training decisions. In particular, the neoclassical production function $F$ must be extended to $2^M - 1$ dimensions (one for each nonempty subset of $M$, representing all the various combinations of assets for which an employee can be trained). Then, the value of $F$ under a given configuration of trained workers will follow from solving a simple linear-programming problem regarding how optimally to allocate the workers over the assets, given their training. Once $F$ is so determined, optimal labor and training decisions follow in the standard manner, by choosing the optimal ray along which to average the extended neoclassical profits in this $(2^M - 1)$-dimensional space.
in how joint surplus should be divided. Groups in control of allocation decisions are likely to assign new positions to retain power, while groups with relatively even power may at times choose to cooperate in order to avoid costly battles between constituencies. External shocks to either the relative productivity or bargaining power of the groups, however, will yield a reallocation of power within the organization. Such a conception is likely to add insight to the tension between joint production and intergroup conflict in organizations. Additionally, it should give some notion of when groups will align with one another and when they will oppose each other, and in what manner coalitions within an organization are likely to form. These questions in turn reflect upon a combination of our present analysis with classical theory of the firm questions; that is, how does intrafirm bargaining affect the scope and organization within a firm?

Finally, it is worth reemphasizing the generality of the framework considered here, and in Stole and Zwiebel (1996), along several dimensions. First, in the companion paper, we demonstrate the robustness of the results that drive all our economic applications to a more general class of bargaining games that allow for any given division of bargaining power, including those which vary with the number of employees. We also demonstrate the correspondence of our bargaining game with the Shapley values of an associated cooperative game. As such, our setting yields a cooperative-game interpretation to the noncooperative bargaining game we have emphasized, and it implies results that are quite resilient to natural changes in the environment.

In a completely different vein, it is also worth noting the wide scope of potential circumstances to which this paper can be applied. While we have discussed the bargaining game in the context of a firm and its employees, the general notion seems just as applicable to any circumstance where holdup confers bargaining power on a collection of agents organized to engage in a mutually beneficial activity. As such, it would seem natural to apply our results to governmental decisions and legislative voting, or to the general question of the formation of social networks.

REFERENCES


