The Excess Sensitivity of Layoffs and Quits to Demand

Robert E. Hall, Stanford University

Edward P. Lazear, University of Chicago and National Bureau of Economic Research

Excessive layoffs in bad times and excessive quits in good times both stem from the same weakness in practical employment arrangements: the specific nature of worker-firm relations creates a situation of bilateral monopoly. Institutions which have arisen to avert the associated inefficiency cannot mimic the separation decisions of a perfect-information, first-best allocation rule. Simple employment rules based on predetermined or indexed wages are in many cases the most desirable among the class of feasible employment arrangements. More complicated contracts which seem to deal more effectively with turnover issues either are infeasible because of informational requirements or create adverse incentives on some other dimension.

I. Introduction

When demand falls, employers lay workers off. When conditions are strong in the outside market, workers quit. Sometimes, the workers who are laid off or quit produce less in their subsequent jobs than they would have produced in their original jobs. Such layoffs and quits are inefficient and undesirable. Practical constraints on the nature of the agreement between employer and worker bias the labor market toward these excessive layoffs and quits. We argue that institutional arrangements have

Partial support for this project was provided by the National Science Foundation.
arisen to eliminate costly bilateral monopoly situations. A by-product of
those arrangements is too many layoffs when a worker’s value is lower
than anticipated and too many quits when a worker’s value is higher than
anticipated.

The idea is a simple one. When workers and firms agree to trade they
are uncertain about some aspects of the value of that trade as well as the
value of their alternatives. After that knowledge is obtained, a bilateral
monopoly situation arises because the value of the match exceeds that of
the next best alternative. At that point, it is costly to decide how the rent
should be split. In anticipation of this difficulty, the terms of the trade
are agreed on so that if trade occurs, it does so according to the previously
specified formula. The same problem arises in many economic relations;¹
in this paper we discuss the way it seems to be handled in practice in the
employment relation.

When the terms of employment are set in advance, not every contin-
geney can be accommodated in a perfectly efficient manner. As we will
show, special provisions to bring efficiency on one margin create incen-
tives for distortions on another.²

A growing literature describes the inefficiencies resulting from em-
ployment arrangements where the employer provides insurance against
income fluctuations.³ The inefficiencies, often characterized as unem-
ployment, arise because contracts cannot simultaneously insure the worker
against business-cycle risk and provide appropriate signals and incentives
for efficient separations.

We wish to deemphasize the insurance motive for three reasons. First,
risk neutrality enormously simplifies the analysis. We can investigate a
number of incentive issues that would be hopeless under risk aversion.
Second, permanent-income theory suggests that business-cycle fluctuat-
ions are only weakly correlated with well-being, so insurance against
the fluctuations may not be of major importance. Third, most of the
variation in earnings is across workers rather than over the lifetime of

¹ See Mortensen (1978) for a clear statement of the general problem. Reder
(1947) is an early example of an examination of how welfare economic implications
change when there is a unique productivity associated with a particular firm-
worker pair.

² This is the point of Lazear (1981). There, in order to induce the worker to
put forth sufficient effort, the earnings relationship is distorted; severance pay
and other institutions arise to provide efficient allocation among all dimensions.

³ The earliest papers were Baily (1974) and Azariadis (1975). More recent papers
in this area are Azariadis (1980), Arnott and Stiglitz (1981), Green (1981), Green
(1979) provides a survey and many additional citations. Some of the recent models
generate underemployment while others generate overemployment. Only Cooper
(1983) examines the case of bilateral asymmetric information, the subject of this
paper. What all these papers have in common is an emphasis on the trade-off
between insurance and efficiency.
individual workers. Though workers are very concerned about the risk associated with their unknown lifetime potential to earn, this risk is insurable because of severe moral hazard. Insured workers are too likely to reduce effort below efficient levels.

Unlike most of the earlier work on labor contracts, we assume from the start that information limitations are bilateral. Layoffs may occur when the employer learns of a disappointment in demand which is not public knowledge and cannot be made a contingency in the employment arrangement. In parallel, quits may occur when workers learn of favorable developments in the outside market, again in a way that is not public knowledge and not a contingency in their contracts. Because an important fraction of all permanent job separations take the form of quits, an analysis with a claim to describing the contemporary American labor market must allow quits to play an active role. We also note that information limitations make the distinction between quits and layoffs highly meaningful; our work departs from the position taken by Becker, Landes, and Michael (1977) that separations always occur when they are to the mutual benefit of the employer and the worker and that the distinction between quits and layoffs is meaningless.

We argue that simple arrangements involving predetermined wages or unilateral wage determination are in widespread use because they perform better in many respects than more complicated contracts. Occasional inefficient separation is a by-product of these simple contracts. Central to our approach is the point that surplus quits and surplus layoffs have the same origin: specification of the terms of trade before all the relevant information is available and the inability to make the terms of employment fully contingent on the information.

Of course, the claim that employment decisions tend to take place with respect to a predetermined wage is hardly a novel one. It is the backbone of a large body of macroeconomic thought and, usually less formally, permeates a good deal of labor economics as well. But the theoretical foundations of theories of employment fluctuations based on predetermined wages have been questioned sharply, especially by the equilibrium school (Barro 1977). We hope to respond to these questions.

We proceed as follows: In Section II we argue that contracts, or at least

---

4 See Lillard and Willis (1978), Lillard and Weiss (1979), and Hall and Mishkin (1982). Altonji and Ashenfelter (1980) show that even with past events treated as fixed effects, most personal wage variation arises from transitions between jobs. Although the absence of variation in individual earnings over time might possibly be attributable to insurance, we find this unlikely because cross-sectional variation dominates time-series variation in earnings among the self-employed as well.

5 With risk neutrality and only a single dimension of unobserved shifts in demand, the contract problem has a simple, efficient solution in which compensation is contingent on the level of employment chosen by the employer. See Hall and Lilien (1979).
informal understandings about employment, have an important role in the labor market. Section III explores a simple model and compares three kinds of contracts. We show that no contract dominates the others; each type brings inefficient separation in some circumstances. Consequently, turnover rates are chronically too high. In Section IV, we examine the performance of the three simple contracts for cases of firm-specific capital and positive correlation between demand and supply surprises. Section V considers more unusual contracts. Though some of them avoid inefficient separations, their requirements for information make them infeasible. Section VI considers some additional issues, and conclusions are presented in Section VII.

II. Why Have Labor Contracts at All?

Even with long-term employment and important specific capital, employment contracts are not a logical necessity. A complete theory of efficient job retention and separation is outlined in footnote 4 of Becker et al.'s paper on divorce (1977). Let the firm and the worker bargain over compensation and work, after the information about product is known to the firm and information about alternative opportunities is known to the worker. With specific capital, the bargaining problem is not a zero-sum game, and its outcome cannot be predicted from any widely accepted theory of bargaining. Still, the parties ought to come to an efficient conclusion somewhere on the contract curve. If they decide to separate when retention is efficient, a further step in bargaining is possible that will make both parties better off by taking advantage of the benefits of continued employment. A similar step should prevent an inefficient retention.

The simple period-by-period bargaining solution is widespread in the labor market. Even when contracts exist, they can be overruled by direct bargaining, and this sometimes happens when developments occur that are totally out of the range contemplated by the contract. As Reider (1947) recognized early, the bilateral monopoly aspects of a labor contract can make direct bargaining an enormously expensive process. Those without confidence in their bargaining skills favor a well-defined employment arrangement where bargaining is not part of the process of wage determination.

As our subsequent discussion will show, the unilateral rights of quit and layoff are an important intrinsic feature of labor contracts; they are not artifacts of legal restrictions on the enforcement of long-term contracts. When conditions in the firm call for a reduction in labor input, a layoff initiated by the employer is very likely the right way to bring about the efficient reallocation of labor. Similarly, when the surprise comes in the outside market for the worker's services, a quit is the natural way to bring about the efficient reallocation.
III. Basic Considerations and Three Simple Contracts

Sometimes it is useful for employer and worker to agree on the terms of trade before all the relevant pieces of information become available. By the time the state of product demand and conditions in the outside labor market become known, a good deal of specific capital has developed in the employment relation. As a consequence, postponement of negotiations to the time when the information is known creates a bargaining situation with bilateral monopoly. Both parties gain privately from clever strategic behavior. But this prisoner’s dilemma can be eliminated to the benefit of both parties by reducing the employment terms to a formula which will be the subject of ready agreement before the job-specific capital is formed.

In our model, a worker and a firm come to an agreement in period 0 about the terms under which work will take place in period 1. No work occurs in period 0, but training and other activities take place to form job-specific capital. To keep the focus on the issue of job separations, we assume that work is a binary choice—either it occurs for a standard number of hours in period 1, or there is a separation and no work occurs. Although we will consider more general arrangements in the next section, most of our points can be made by considering three simple contracts.²

Contract 1 (predetermined wage): A wage is agreed upon in period 0 and work occurs in period 1 at that wage unless one side opts for no work through a layoff or quit.

Contract 2 (firm sets wages): The firm announces a unilateral wage offer in period 1 and the worker chooses to work at that wage or not at all.

Contract 3 (worker sets wage): The worker announces a wage demand in period 1 and the firm chooses to employ the worker at that wage or not at all.

For the reasons discussed in the Introduction, we assume both the firm and the worker are risk neutral (risk aversion merely reinforces the conclusions of this section). Denote the worker’s marginal product at the firm by $M$ and the value of the alternative use of the worker’s time (at another firm or at home) by $A$. The worker privately observes $A$ and the firm privately observes $M$ at the start of period 1. Information is bilaterally asymmetric.

We take as the best contract the one maximizing the sum of expected values to both parties. We allow a lump-sum, nonallocative payment from one party to the other to achieve a mutually satisfactory distribution of

² Similar contracts are analyzed in Weitzman (1981), but he does not consider a no-trade clause in the contract with predetermined price. The role of no-trade provisions is the essence of our discussion. Weitzman also lets the quantity traded be a continuous variable.
the benefits of the bargain. Maximizing the sum of expected values is equivalent to maximization of profit subject to a minimum expected utility for the worker and equivalent to maximization of utility subject to a minimum profit for the firm. If \( X \) is a random variable equal to one if work occurs and zero if not, the goal is to maximize

\[
E[XM + (1 - X)A].
\]

Contracts attempt to make \( X = 1 \) when \( M \) is large and \( X = 0 \) when \( A \) is large; that is, to have work take place when productivity in the firm is favorable and to have a separation when the outside market is favorable. The first-best contract sets \( X = 1 \) when \( M \) exceeds \( A \) and \( X = 0 \) otherwise. Other contracts will be judged against this basic efficiency criterion that work should occur when and only when the worker has a comparative advantage in the firm \( (M > A) \).

In general, contracts provide for a certain amount of compensation to be paid at the outset and then specify additional compensation after information becomes available in period 1. The initial compensation is effectively a lump sum and has no role in determining whether employment occurs; it is available to offset any distributional effects the later provisions may have. In a contract of type 1, a payment, \( S \), flows from the firm to the worker at the outset and then a fixed wage, \( W \), is paid if work takes place in period 1. After the contract is signed, the relevant marginal payment is \( W \) so all decisions depend only upon it.

In contract 1, employment is determined in the following way: The firm learns \( M \) at the beginning of period 1. It lays the worker off if \( M \) is less than \( W \). The worker learns \( A \) at the same time and quits if \( A \) exceeds \( W \). Of the four possible outcomes, work occurs in only one, when there is neither a layoff nor a quit. The operation of the contract is displayed in figure 1, which borrows from Hashimoto and Yu (1980) and Carmichael (1981). Whenever the realization of \((M, A)\) lies below the 45-degree line in figure 1, it is efficient for work to occur. But the firm will lay the worker off in the whole region to the left of the vertical line \( M = W \), and the worker will quit in the whole region above the horizontal line \( A = W \). The two shaded triangles describe the potential inefficiencies from the predetermined wage contract.

In the lower-left triangle, an inefficient layoff occurs. Conditions are poor in the firm (\( M \) is low) but even worse in the outside market (\( A \) is even lower). Though the worker has a comparative advantage in the job, the contract fails to make the employer take that comparative advantage into account. In the upper-right shaded triangle, conditions are good in the outside market (\( A \) is high) but are even better in the firm (\( M \) is even higher). Again, the worker has a comparative advantage in the current job, but in this case the contract fails to provide the worker the appropriate incentives to remain and a quit occurs.
Users of contract 1 will choose $W$ in advance to maximize the joint return as given in equation (1), subject to a distribution of $X$ in which it has value only if $A$ is less than $W$ and $M$ is greater than $W$.

Contracts of types 2 and 3 try to make use of information that becomes available in period 1. In contract 2, the firm observes $M$ before calling out the wage offer, and in contract 3, the worker observes $A$ before calling out the wage demand. These procedures add some flexibility which may reduce the loss from inefficient separation. But the gains are countered by the inefficiencies brought by the exercise of monopoly power by the party responsible for setting the wage. The private benefit considered as a function of the wage set at the beginning of period 1 looks quite different from the joint benefit, so efficient wage-employment behavior is impossible in contracts 2 and 3. Still, the added responsiveness of contracts 2 and 3 can make them superior to the fixed-wage contract 1 in some circumstances.

Consider contract 2. The noncontingent payment, $S$, is agreed on in advance, and the firm is free to choose $W$ after having observed $M$. Both the worker and the firm recognize that the firm will ignore the benefits from the match which do not accrue to the firm. Even so, there may be a value of $S$ such that both the firm and the worker are better off under contract 2 than any other feasible contract.

The firm selects the wage in period 1 after observing $M$ so as to max-

![Graph](image.png)

**Fig. 1.**—Outcomes under the fixed-wage contract
imize \( E[X(W)(M - W)] \), where \( X(W) \) is the stochastic labor supply schedule of the worker; \( X(W) = 1 \) if \( W \) is not lower than \( A \) and is zero otherwise. For expositional simplicity, we will assume for the moment that \( M \) and \( A \) are independent, so we can write the cumulative distribution function of \( A \) as \( G(A) \), the probability that the value of outside opportunities is no greater than \( A \). Because \( M \) and \( W \) are nonrandom from the point of view of the firm in period 1, the expected profit is just \( G(W)(M - W) \). Expected profit reaches its maximum when \( W = M - G(W)/g(W) \). We conclude that the firm will always call out a wage which is less than the worker's marginal product. The depressed wage offer creates the possibility of a quit when the firm would have been willing to offer the worker a wage sufficient to attract him to remain on the job. As a monopsonist (ex post), the firm shades the wage downward because there are many potential values of \( A \) which lie below the observed \( M \) and the firm reaps profit in these cases by offering the worker a wage less than his full value to the firm.

Contract 3 is similar. Here the worker is given the power unilaterally to set a wage demand in period 1. Both the firm and the worker know that the worker will act as a monopolist in period 1, but the added flexibility may make this arrangement superior. The worker chooses the wage by maximizing \( E[X(W)W + [1 - X(W)]A] \) after he learns \( A \). Now the first-order condition for maximum expected earnings is \( W = A + [1 - F(W)]f(W) \), where \( f(M) \) is the density of \( M \) and \( F(M) \) is its cumulative distribution. As a monopolist, the worker always sets a wage above the value of the alternative use of time, even though this will sometimes result in inefficient separation.

As one might expect, the choice among contracts 1, 2, and 3 depends upon the joint distribution of \( M \) and \( A \). No single contract type is superior under all circumstances. We prove this by counterexamples shortly. But the choice of contract type is not the main focus of our work. Rather, we want to say something about the relationship between demand fluctuation and job separations. We now have the necessary machinery to proceed.

To provide a feel for the forces at work here, we analyze a special case. Assume that \( M \) and \( A \) are independently and uniformly distributed between zero and one. Then the analysis earlier in this section shows that the best predetermined wage is \( W = 1/2 \) and the expected loss from inefficient separations is \( 1/24 \). Under contract 2, the wage policy maximizing expected profit for firms is \( W = M/2 \), because \( g = 1 \) and \( G(W) = W \). The loss from inefficient separations is the area of triangle \( ABC \) in figure 2, calculated as \( E[X^*(M, A) - X(M, A)][M - A] \). Here \( X^*(M, A) \) is the first-best employment rule: \( X^* = 1 \) if \( A \) is no greater than \( M \) and zero otherwise; and \( X(M, A) \) is the employment rule when the employer sets the wage: \( X = 1 \) if \( A \) is no greater than \( M/2 \) and zero otherwise. The expected loss is also \( 1/24 \).
Under contract 3, where the worker makes a wage demand after learning $A$, the best policy is to ask for $W = A/2 + 1/2$, because $f = 1$ and $F(W) = W$. The expected loss from inefficient layoffs is the area of triangle $ABD$ in figure 2 and is computed from the expression just given with $X(M, A) = 1$ if $M$ is not less than $A/2 + 1/2$ and zero otherwise. Again, the expected loss is $1/24$.

In all three contracts, the value of the lump-sum component of compensation, $S$, can be set to provide the appropriate distribution of expected profit and earnings to the two parties.

In this example, the three contracts yield identical joint expected losses and are equivalent in that sense, even though they are very different in actual outcome and in the interpretation given by the two parties to events. In the predetermined wage contract, there are both quits and layoffs. In the firm-sets-wage contract, there are only quits. In the worker-sets-wage contract, there are only layoffs. No one contract is preferred to another without bringing in other considerations. We shall examine other criteria for contract choice in a later section, but before doing so, we state our main point:

Too many quits occur when conditions in the market are good but conditions in the firm are even better. Too many layoffs occur when conditions are poor in the firm but even worse in the market. “Too many” or “inefficient” is defined relative to the perfect information, first-best optimum.

First, consider the predetermined wage contract illustrated in figure 1. In our example, the optimal predetermined wage is $W = 1/2$. When demand is relatively weak ($M$ is below $1/2$), the worker is laid off. Sometimes these layoffs are efficient (whenever $M < A$). At other times, it would be in the joint interest of the two parties to retain the match because

![Diagram](image-url)

**Fig. 2.**—Outcomes under firm-sets-wage and worker-sets-wage contracts
the value of the worker's alternative use of time is below his value to the
firm, even though that value lies below the wage rate. If worker and firm
could renegotiate the terms of trade, work would take place. But it is
exactly the costly renegotiation that the fixed-wage contract sought to
avoid. The price paid is an occasional inefficient layoff. The reverse is
due during good periods for the firm, when $M$ exceeds $1/2$. The firm
never lays the worker off, but the worker will quit whenever $A$ exceeds
$1/2$. In the upper shaded region in figure 1, the worker quits when staying
on the job would have been mutually beneficial. Thus, workers quit too
frequently when conditions are favorable and firms lay workers off too
too frequently when they are unfavorable. Employment is excessively sen-
sitive to fluctuations in demand.

In this simplest case, it is interesting to note that a given layoff is more
likely to be inefficient when $M$ is only a little below the wage, $W$. Because
the layoff decision is insensitive to the value of $A$, the higher is $M$ the
more likely it is to exceed $A$.

Next consider contract 2, where the firm makes a wage offer after
having observed $M$. There can be no layoffs with this contract, but in-
efficient quits are correspondingly more likely. With reference to figure
2, it is clear that a worker opts to quit whenever $A$ exceeds $W$, that is,
everywhere above the line $AC$ in figure 2. A worker is almost certain to
quit when $M$ and therefore $W$ are low. Further, since $M$ is low, most
quits are efficient in that the worker’s alternatives exceed his value to the
firm. For higher $M$, however, the probability of a quit is lower (the
probability is one at $M = 0$ and $1/2$ at $M = 1$). But the incidence of
inefficient quits is higher for higher values of $M$. The triangle of inefficient
quits, $ABC$ in figure 2, becomes taller on the right; that is, $E(M - A)$
($X^* - X$) is an increasing function of $M$. Inefficient quits are more likely
to occur when the firm experiences good times relative to the rest of the
market, even though the number of quits actually falls during these pe-
riods. The inefficiency stems from the firm’s extraction of monopsony
rents from workers by setting the wage below $M$.

Finally, consider contract 3, where the worker makes a wage demand
after observing $A$. Quits do not occur, but layoffs are correspondingly
more of a problem. The wage demand always exceeds $1/2$, so a layoff is
certain if $M$ is below $1/2$. Here, as in the fixed-wage case, layoffs which
occur when $M$ is just below $1/2$ are more likely to be inefficient than
those which occur when $M$ is well below $1/2$ because the probability that
$A$ exceeds $M$ rises as $M$ falls. For values of $M$ above $1/2$, the closer $M$ is
to one, the lower is the probability of a layoff and the lower is the
probability of an inefficient layoff.

To summarize this section, inefficient layoffs occur when the value of
the worker to the firm is low, but the worker’s alternative opportunities
in the outside labor market are even weaker. With a predetermined wage,
the firm reacts only to the relationship between the wage and the internal value of the worker and cannot moderate layoffs when opportunities are poor in the outside market. Letting the worker set the wage does not reduce the incidence of inefficient layoffs in bad times by lowering the wage—on the contrary, the worker may be so aggressive in exerting monopoly power as to set a wage that guarantees layoff whenever demand facing the firm is below average, no matter what is happening in the labor market. Similarly, inefficient quits are frequent because the worker reacts only to the relationship between the wage and the alternatives in the labor market. With a fixed wage, the worker rationally ignores the possibility that things are good elsewhere but even better at the firm. Again, letting the firm set the wage might help solve this problem, but in our example, the firm exploits its monopsony power so aggressively that it never offers a wage high enough to forestall an inefficient quit.

The shortcomings of the three simple contracts considered in this section suggest there is a role for richer types of contracts. Later in the paper, we consider a wider set, but we conclude that no practical contract solves all problems effectively. Excess sensitivity of separations to demand seems an inescapable feature of the labor market.

IV. More on the Three Simple Contracts

In the example of the previous section, all three contracts had the same expected joint loss from inefficient separation. Our choice of a joint distribution for $M$ and $A$ explains this coincidental tie. In this section, we show how the characteristics of the joint distribution favor one contract over another. We look at the case of a mean of $M$ well above the mean of $A$, which might arise because specific capital gives the worker a strong comparative advantage in the current job. We also examine the case of positive correlation between $M$ and $A$, a likely case for business-cycle applications.

In discussing the issue of comparative advantage, we will retain our assumption of independence of $M$ and $A$. Suppose, in our earlier example, that the worker obtains one unit of firm-specific human capital in period 0, so that the distribution of $M$ in period 1 is uniform between 1 and 2. At the same time, nothing happens to the distribution of $A$, because the investment improves the worker’s productivity at the current firm only. The situation is illustrated in figure 3. The best fixed-wage contract sets $W = 1$ so that neither quits nor layoffs ever occur. Because $M$ always exceeds $A$, separations are never efficient and the fixed-wage contract is first-best. Contract 2 still makes firms offer a wage of $M/2$, which brings inefficient quits when $(M, A)$ realizations fall in the triangle $ABC$. Contract 3 makes workers demand a wage of $A/2 + 1$, which brings inefficient layoffs when the $(M, A)$ realizations lie in triangle $DBE$. Obviously the fixed-wage contract is superior to either of the alternatives. Specific capital
makes it more likely that work should occur. A contract with a fixed wage, above most of the alternatives but below the likely marginal product at the firm, will give close to the efficient pattern of separations. As a general matter, we conclude that specific capital and other conditions leading to strong comparative advantage to the worker in the current job favor the fixed-wage contract over the two variable-wage contracts.

This reasoning has implications for variability in wages and separations over the life cycle. Young workers have less specific capital, so they are less likely to have a fixed-wage contract. Accordingly, compared to their older colleagues, their wages and employment should vary more over the business cycle. We cannot make any similar statements about variations in wages across workers because there is so much heterogeneity, which varies with age.

Positive correlation between $M$ and $A$ works in the opposite direction, improving the efficiency of the flexible wage contracts relative to the predetermined wage contract. To see this, consider the case of perfect correlation between $M$ and $A$. When the firm learns its $M$, it also learns $A$. It faces a labor supply function that is perfectly elastic at the wage $A$, so it has no monopsony power and simply offers $A$ when $A$ does not exceed $M$. Inefficient separation never occurs.\footnote{Weitzman (1981) presents an analogous result.}

V. Other Contracts and Other Criteria

We have restricted our attention to three types of wage arrangements because we believe that these are the most prevalent contract types. We

![Diagram](image-url)

**Fig. 3.**—Outcomes when $M$ always exceeds $A$
have also cast most of the discussion in terms of separation efficiency and have ignored other efficiency criteria by which one may evaluate contracts. In this section, we briefly consider a wider class of contracts and a number of criteria by which to choose among them.

In this wider class, the employer makes a wage offer, $W_O$, after learning $M$, and the worker makes a wage demand, $W_D$, after learning $A$. Two mediation formulas translate the offer and the demand into a wage paid, $W_P$, and a wage received, $W_R$:

$$W_P = W_R(W_O, W_D)$$

$$W_R = W_R(W_O, W_D).$$

In addition, there is a lump-sum payment, $S$. We investigate the features of arrangements where wages received are not necessarily the same as wages paid, and a third party makes up the difference. Though such arrangements have some very desirable properties, we conclude that the involvement of third parties is impractical.

In the wider class of contracts, severance pay may be part of the compensation plan, and the amount of severance pay may be one amount, $Q$, if the worker quits and a different amount, $L$, if the employer lays the worker off. An even more general contract would permit the $L$ and $Q$ received to differ from the $L$ and $Q$ paid, as investigated by Carmichael (1981), but we do not pursue that generalization.

We evaluate the members of the wider class by the following criteria:

1. Information feasibility and efficiency: The parties have the necessary information, resources, and the appropriate incentives to carry out the terms of the contract. The contract does not create incentives to expend resources generating false information. Further, the stochastic appearance of additional information does not render the contract inefficient.

2. Separation efficiency: Work occurs when, and only when, $A$ is not greater than $M$. This was the single criterion considered in the earlier sections.

We also note there is an issue of investment efficiency. Efficient investment in job search and general and specific human capital occurs if and only if the joint return to the worker and firm exceeds the joint costs. But none of the contracts we consider satisfies investment efficiency. Because investment issues are a complex topic by themselves, we defer discussion to a later paper.

We can summarize our findings about the three simple contracts with respect to these two criteria in the following compact way: All three are feasible and efficient in their treatment of information, but none provides separation efficiency. Now we will turn to contracts that promise separation efficiency but are impractical from the point of view of information.
Piece Rates

First is a piece-rate contract. The firm pays the worker $M$ and the worker decides whether to work for the firm or not. In our notation, $W_p = W_R$ and $L = Q = 0$. By rewarding the worker fully for what he produces for the firm, the worker is made to internalize the separation decision, and this contract brings full separation efficiency. But the piece-rate contract is not feasible from the point of view of information. The worker cannot verify that the rate of pay actually equals his productivity. Only firms know $M$, but they have no incentive to reveal it truthfully. Further, firms face an incentive to devote costly resources to concealing or falsifying information about $M$. If the worker simply accepts the firm’s announcement of $M$ without any verification, the purported piece-rate contract is exactly the same as the firm-sets-wage contract studied earlier in the paper, where separation efficiency fails.

Market Wage

A related contract pays the worker the market wage, that is, $W_p = W_R = A$ and $L = Q = 0$ (with $S$ chosen as usual to distribute the rents). Now the problem is the firm’s inability to verify the worker’s claim about the opportunities in the outside labor market, $A$. The worker faces incentives to produce evidence of highly favorable outside conditions, an activity which is costly or which effectively converts the contract into the worker-sets-wage form. Either way, efficiency fails.

The Expectations Principle

A third type of contract amounts to treating an employment contract as an ordinary commercial contract under the common law. It has the same provisions as the predetermined-wage contract treated earlier, but instead of walking away from the employment relationship without further financial consequences, the party that dissolves the relationship must compensate the other for losses inflicted. Under the common law, the departing party must compensate the other by the amount of the expected gain evaluated at the time the departure (or breach) occurs. Let $W$ be the agreed wage. If the firm lays the worker off, it must compensate the worker for the difference between the contract wage, $W$, and the actual value of the worker’s time, $A$: $L = W - A$. A quitting worker must compensate the firm for the difference between marginal product and contract wage: $Q = -(M - W)$; recall that $Q$ is a payment from firm to worker, so it is negative in this case. Again, there is a lump-sum payment, $S$, to distribute rents without any allocational consequences.

A familiar result from the economic theory of contracts establishes separation efficiency for this contract. The firm makes its layoff decision

\[ \text{\textsuperscript{8}} \text{See Shavell (1980, 1981) and Polinsky (1981).} \]
by comparing profit from employment, $M - W$, to profit with a layoff, $-L = -(W - A)$. The firm will choose employment if and only if $M - W$ is not less than $-W + A$, that is, if and only if $M$ is not less than $A$, our original efficiency condition. Similarly, the worker compares earnings on the job, $W$, to earnings in the case of quit, $A - (M - W)$. Again, the worker chooses to quit if and only if $A$ exceeds $M$, as required for separation efficiency.

Under the expectations principle, there are situations when the worker wants to quit and the firm simultaneously wants to lay the worker off. Any point in the northwest quadrant of figure 1 has this character. The firm stands to gain if the resulting separation is labeled a layoff rather than a quit—profit from a layoff is $-(W - A)$, whereas profit from a quit is $M - W$, which is smaller because $M$ is less than $A$. Similar logic shows that the worker prefers that a separation be labeled a quit. The two parties may spend resources trying to be the first to bring about a separation. But this problem exists with every commercial contract and does not seem to be a major difficulty.

Although the expectations principle brings separation efficiency, it does not satisfy our requirements with respect to information. The firm does not know how large a payment it will be obligated to make to the worker in the event of a layoff, so it will not make an efficient decision, in general. The same holds for the worker. Each party faces an incentive for costly research to acquire the other side’s information. Further, because compensation is effectively contingent upon the values of $A$ and $M$, each side has an incentive to expend resources to make the apparent values of $A$ and $M$ differ from the actual values. Finally, the expectations principle sets up the wrong incentives for investment in period 0, an issue known in the legal literature as reliance. Again, the complexity of investment issues prevents us from pursuing the question in this paper.

Offer Matching

A closely related contract involves offer matching.\(^9\) Again, a wage is set in advance, and a lump-sum payment, $S$, distributes rents, but actual compensation is raised if $A$ exceeds the wage. The worker has an incentive to bring concrete evidence to the firm about the value of $A$ in the form of a job offer. If $A$ exceeds $W$ but still falls short of $M$, the firm raises the wage to $A$. If $A$ exceeds $M$, the worker quits, efficiently. One-sided offer matching of this kind eliminates inefficient quits but does not limit inefficient layoffs.

To prevent inefficient layoffs, when $M$ falls below $W$ but exceeds $A$ a more dubious form of offer matching is required, in which the worker agrees to accept the lower wage when the firm produces an offer from a

\(^9\) See Mortensen (1978) and Diamond and Maskin (1979).
worker to work at the lower wage. Overwhelming informational obstacles limit this procedure. If a single offer to work at lower wages is enough to permit cutting the wages of a number of existing workers, the firm has an incentive to make a side arrangement with somebody to make a fraudulent offer. At the extreme, workers are powerless to verify purported offers from others to work at lower wages. Then the offer-matching contract becomes the firm-sets-wage contract. Furthermore, the evaluation of offers becomes impractical when nonwage dimensions of jobs are important. Lastly, verification that offers are genuine on either side is costly and difficult.

Still, one-sided matching of offers received by workers is an important feature of the labor market and probably makes a contribution to reduction of inefficient quits. Offer matching does not stimulate excess investment in job search. A worker searches if the expected return from search, \( A - W \), exceeds the cost. Offer matching does not change that. If the firm does not match the offer, the worker leaves and receives \( A \). If the firm does match the offer, the worker stays and receives \( A \). Costs and returns to the worker are the same. But offer matching does encourage fraud, and the efficient employment arrangement may prohibit responding to offers in order to eliminate investment in phony offers. Under offer matching, a worker may arrange with a third party for a fraudulent offer and split the resulting wage increase. Even if an outside offer is genuine, the worker may engage in a costly and potentially inefficient game of presenting the offer to his employer for matching even though he would rather stay on the current job at the current rate of pay. Nonpecuniary dimensions of jobs are difficult for the employer to verify.

The Bilateral Vickrey Contract

This contract applies William Vickrey's (1961) auction principle to both sides of the employment arrangement.\(^\text{10}\) The bilateral Vickrey contract involves a third party because wages paid always fall short of wages received. The employer makes a wage offer and the worker makes a wage demand; each acts at the same moment, without knowing what the other has announced. If the offer is at least as good as the demand, employment occurs, but the employer pays the demanded wage while the worker receives the offered wage. In our notation, \( W_R = W_O \) and \( W_P = W_D \). In addition, the employer and the worker pay the third party an amount equal to the expected value of \( W_P - W_R \) (with the convention that \( W_P - W_R = 0 \) when a separation occurs), and \( S \) is paid by the worker to the firm as a lump sum.

In this setup, the incentives induce the employer to set its wage offer equal to the marginal product, \( M \), and also induce the worker to set the

\(^{10}\) Joseph Stiglitz assures us that this application of Vickrey's idea is "well known in the principal-agent literature," but we have not found a written discussion.
wage demand equal to the value of the alternative use of time, $A$. From the employer’s point of view, the wage offer has no influence on the wage cost of the worker; it only controls whether employment occurs. A wage offer above $M$ would create the possibility of employment at a wage in excess of $M$, and results in a loss with no compensating gain in other states. A wage offer below $M$ does not save the firm any money; it deprives the firm of the possibility of profitable employment when the worker’s wage demand is below $M$. Consequently, the firm always sets its wage offer to $M$. Similar logic shows that the worker always demands $A$. Work occurs if and only if $M$ does not fall short of $A$, exactly the condition for efficiency.

The role of the third party in the bilateral Vickrey contract creates serious problems and is presumably the reason that such contracts are never found in practice. Collaboration between the worker and employer can victimize the third party, so the supply of willing third parties is limited. Worker and employer maximize joint benefits by making very high wage offers and very low wage demands. In this respect, the incentives to reveal the true $M$ and $A$ in the bilateral Vickrey contract are an illusion. The third party would have to try to verify $M$ and $A$, which is costly in itself and creates incentives for the firm and worker to expend resources establishing false values of $M$ and $A$.

Coordinated Severance Pay

In the firm-sets-wage contract studied early in the paper, the firm always sets a wage that is too low because of the firm’s monopsony power. If the contract embodies a wage subsidy, the firm can be induced to set the right wage and bring efficient separations through quits. Whether such a scheme can bring exact efficiency depends on the information available at the time the wage decision is made relative to the information available at the time the contract is signed. In the extreme case where the firm learns nothing about the likely value of $A$ after contract signing, the analysis of d’Aspremont and Gerard-Varet (1979) can be applied to this problem.¹¹ Let $Y$ be the increment to compensation associated with working, so compensation in the case of work is $W = Y + Q$, and let the contract embody the formula, $Q(Y)$, which assigns a level of quit pay given the firm’s decision about $Y$. The parties maximize joint benefits by choosing a $Q(Y)$ that brings efficient quits by exactly offsetting the monopsony influence that otherwise tends to make $Y$ too low and causes excess quits.

When $Q(Y)$ is properly chosen, the firm’s profit-maximizing $Y$ will be exactly its $M$—when efficient separations are achieved entirely through quits, the extra compensation for working, $Y$, must equal the worker’s

¹¹ We thank Oliver Hart for suggesting this line of attack and for pointing out the reference.
marginal product, $M$. The right $Q(Y)$ will make the firm spontaneously reveal the true value of $M$. When the time comes to choose $Y$, the firm will try to maximize expected profits, $E[(M - Y - Q(Y)) X - Q(Y)(1 - X)]$. As before, $X$ has the value of one if work occurs and zero for a quit. If $G(A)$ is the cumulative distribution of $A$, then the probability that work occurs when the firm announces an incremental wage of $Y$ is $G(Y)$, and expected profit is $(M - Y)G(Y) - Q(Y)$. Profit reaches its maximum at $Y = M - [G(Y) + Q'(Y)g(Y)]/g(Y)$; this is just our earlier expression for the firm-sets-wage case with the subsidy term $-Q'(Y)/g(Y)$, added. In the special case where $G(Y)$ is known at contract time, the optimal subsidy formula is simple: Let $Q'(Y) = -G(Y)$. Then the subsidy exactly cancels the monopsony term, $-G(Y)/g(Y)$, and the firm is induced to announce the efficient incremental wage, $Y = M$.

In this special case, the piece-rate approach to compensation is feasible even though the worker cannot verify $M$. The tax or penalty embodied in the quit pay is just enough to induce the firm to set the right piece rate voluntarily.

In general, it will not be possible for the contract to anticipate the monopsony power of the firm. If the firm learns something about the state of the labor market after the contract is signed but before setting the wage, then the relevant $G(Y)$ cannot be written into the contract. The distribution of $A$ conditional on information available at contract time is not the relevant distribution; if the contract sets a subsidy based on that distribution, it could turn out to be inefficiently high. The subsidy is proportional to the probability of work. If the probability was thought to be high at contract time for some particular value of $Y$, and that $Y$ turns out to be the optimal one for the firm to choose, but then the actual probability of work is low, the subsidy could be much too high. Inefficient retentions cannot be avoided if new information becomes available to the firm. As such, this contract violates information efficiency because the stochastic appearance of additional information creates an inefficiency.

Though coordinated severance pay cannot solve the separation problem exactly except under highly unrealistic circumstances, it does point to the desirability of a subsidy formula in the firm-sets-wage contract. The firm always has some monopsony power, so some degree of subsidy to the wage is desirable to offset the power. In cases where unilateral wage determination is the preferred solution, we would expect to find contract provisions or implicit understandings that lower pay for work will be accompanied by higher severance pay or other elements of compensation not related to the amount of work.

---

12 This kind of contract is a variant of a Groves (1973) mechanism for inducing truthful revelation. Most discussions have one party announce the value of a variable and let the contract translate the announced value into a price. In the labor market, it seems to us more natural for the firm to announce a wage directly. Analytically, the two procedures are identical.
Without going into the details, we note there exists a formula relating layoff compensation to the wage announced by the worker which induces the worker to reveal his true $A$ and thereby bring about separation efficiency. The same defects attend this technique as the symmetric one for employers just discussed.

Following is a brief summary of our conclusions about all the contracts treated in the paper:

1. The fixed-wage, firm-sets-wage, and worker-sets-wage contracts are all feasible with respect to information but fail separation efficiency.

2. Piece rates, market wages, and the expectation principle bring about separation efficiency but require that both parties possess information which in many cases is private to one side.

3. One-sided offer matching eliminates inefficient quits, while two-sided offer matching eliminates all inefficient separations. But offer matching stimulates fraudulent offers.

4. The bilateral Vickrey contract brings about separation efficiency but places unreasonable informational requirements on a third party and invites collusion between the third party and the worker or the firm.

5. Coordinated severance pay is both feasible and consistent with separation efficiency as long as the firm learns nothing new about outside opportunities for the worker between the framing of the contract and the setting of the wage. The arrival of new information will bring a violation of separation efficiency.

Probably the most important message of this section is the absence of a dominant contract. Even without risk aversion, most arrangements fail to satisfy important criteria. In particular, none of the contracts that achieve separation efficiency come to grips with bilateral limitations on information. In our view, the information criterion comes first in the ranking of contracts. The piece-rate, market-wage, and expectations-principle contracts are often infeasible from the start because of the insuperable obstacles to direct measurement of $M$ and $A$ to the satisfaction of both parties. Offer matching and the bilateral Vickrey contracts try to induce truthful revelation, but they create opportunities for collusion that render them impractical. Coordinated severance pay has the opposite problem—it fails when the firm has more information than anticipated by the contract. Consequently, we reject all of the contracts that claim to achieve separation efficiency. We conclude that excess layoffs and quits are a necessary consequence of institutional arrangements which are the best solution to informational inadequacies.

VI. Other Issues
Penalties for Quits and Layoffs

One of the extensions considered in Section IV can be applied to the simple fixed-wage contract considered earlier in the paper, namely, the provision of different penalties for a separation depending on who initiates
it. It may be possible to improve the separation efficiency of a fixed-wage contract by allowing $L$ to differ from $Q$. The simple fixed-wage contract brings inefficient separations but never inefficient retentions. With $L$ different from $Q$, the two sources of inefficiency can be traded off against one another.\footnote{This is the essence of the argument of Green and Kahn (1981) and Cooper (1983) that overemployment as well as underemployment can occur under optimal second-best contracts. Green and Kahn obtain the result under unilateral asymmetric information and risk aversion. Cooper allows for bilateral asymmetric information, but also emphasizes risk aversion.} Consider a contract which pays $W$ if work occurs, $Q$ if a quit occurs, and $L$ if a layoff occurs. The firm lays the worker off if $M < W - L$. The worker quits if $A + Q > W$. Figure 4 illustrates the resulting situation. In the triangle $ABC$, inefficient work occurs. A layoff should occur, but the firm is unwilling because of the layoff penalty. The height of the triangle, $AB$, is $L - Q$, which is equal to its width, $BC$. What indexes the trade-off between inefficient retentions and inefficient separations is not the level of either $Q$ or $L$, but the difference between them. A trade-off takes place only when the layoff pay exceeds the quit pay.

Layoff pay in excess of quit pay creates adverse incentives. The worker who finds $A > W - Q$ can do better if he can induce the firm to lay him off rather than quitting; he will earn $A + L$ rather than $A + Q$. Similarly, when the firm discovers that it would be profitable to lay a worker off, it would prefer that the worker quit. The worker, finding a good opportunity in the outside market, has an incentive to shirk in order to induce a layoff. The firm, finding the worker redundant, has an in-

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure4.png}
\caption{Outcomes with quit pay $Q$ and layoff pay $L$}
\end{figure}
centive to make his life miserable to induce him to quit. Or, even if neither of these responses occurs, the worker and the firm may reach a standoff, where both recognize that a separation is timely but each hopes the other will go first; this can happen anywhere in the northeast quadrant of figure 4. Carmichael (1981) notes that this class of problems can be circumvented by diverting layoff pay to a third party, so that the firm faces incentives to limit layoffs yet workers do not have incentives to stimulate layoffs.

There are no similar obstacles to quit pay in excess of layoff pay, but, as figure 5 shows, contracts of this type are perverse with respect to separation efficiency. The shaded area $ADEC$ of inefficient separations can be eliminated by lowering $Q$ to $L$ with no corresponding increase in other inefficiency. We conclude that $L$ will never be less than $Q$ and will not be too far above $Q$. Lazear (1983) shows that pension benefits are sometimes higher for workers who retire at the firm’s request rather than at their own initiative, but these differences are small.

Unilateral Asymmetric Information

When workers are risk neutral, it should be clear that unilateral asymmetric information allows achievement of the first-best allocation of labor.\(^{14}\) This well-known result is worth restating in the present context.

If one side has all the relevant information, then it is efficient to allow that side to determine the wage offer and to allow the other side to decide on separation. For example, if the firm knows both $A$ and $M$, while the

---

worker knows only $A$, the firm should be given the right to select the wage, $W$. Because the firm knows $A$, it will always offer $A$ when $M$ exceeds $A$ and zero otherwise. The worker will always work when $W$ is at least as high as $A$, so work occurs whenever $M$ is at least as high as $A$. This brings the efficient allocation of the worker’s time. The two parties can agree on a lump-sum component of compensation to insure the desired distribution of profits and utility.

Indexed Contracts

If both parties to an employment contract observe a variable that is correlated with $A$ or $M$, it may be possible to improve the performance of the fixed-wage contract. But the same circumstances also favor the firm-sets-wage or the worker-sets-wage contracts, so the role of indexing is circumscribed. For example, consider an extreme case where the unemployment rate conveys full information about $A$. A contract with a wage indexed to the unemployment rate would provide full separation efficiency, but so would a contract which assigns to the firm the right to choose the wage.

The goal of indexation is to make the various provisions of a contract vary as conditions change so that employment and separation take place in accord with the efficiency condition. To keep the discussion simple, we will make the strong assumption that compensation is exactly the same in the case of a quit as in the case of a layoff. For the purposes of guiding quits and layoffs, there is nothing to be gained by indexing severance pay as well as the wage itself, because both the layoff and quit decisions depend only on the difference between the wage and severance pay. Predetermined severance pay can be combined with the lump-sum component of compensation, so $W$ is the amount of additional compensation paid in the case of work, and no additional compensation is paid in the case of separation.

A wage-indexing provision achieves separation efficiency if and only if $A \leq W \leq M$ whenever $A \leq M$. No matter what wage is set by the contract, a separation occurs if it is efficient. The trick is to prevent inefficient separations by keeping the wage between $A$ and $M$ in those cases where there is room between $A$ and $M$. If there are imperfect, publicly known indicators of both $A$ and $M$, the index formula will give some weight to both, to minimize the probability of violating the efficiency condition.¹⁵

¹⁵ An early discussion of optimal wage indexation, under somewhat restrictive assumptions about the nature of the employment contract, appears in Gray (1976). More recently, Card (1982) has studied the nature of optimal indexing in a multiperiod contract with bilaterally asymmetric information but ignores the issue of separation efficiency.
With respect to $M$, one natural indicator is the price of the firm's product. In the nineteenth century, British coal miners received wages indexed to the price of coal in an arrangement called the "sliding scale." We do not know of any contemporary examples of indexation to productspecific price data. Profit sharing is a closely related type of indexation and is widespread, especially in Japan. As a general matter, indexation to indicators of $M$ puts the responsibility for making separation decisions on the worker. Exact indexation to $M$, which is the piece-rate contract discussed earlier, makes the firm completely indifferent about the level of employment.

With respect to $A$, indicators of conditions in local labor markets are the natural choice. Private wage surveys and government wage indexes appear to have an important role in the wage-setting process, but we are unaware of any instances of formal indexation to outside wage indexes. For the other dimension of labor market conditions, the cost of finding new work, indexation to measures of unemployment would appear to be useful, but again we are unaware of any formal indexation of this type.

VII. Concluding Remarks

No single contract has emerged from our study as a complete, practical solution to the basic problem of deciding whether the efficient use of a worker's time is at the firm or in the outside market. We have discussed the virtues and defects of a number of rules which set compensation so as to bring more or less desirable unilateral decisions by firms and workers about separations. Some of these rules are rejected at the outset for requiring an impractical amount of information. Others are rejected at the next stage for creating serious adverse incentives. None of the survivors brings full separation efficiency under realistic assumptions about information.

With practical labor contracts, inefficient separations will occur in a characteristic way. Layoffs and quits are excessively sensitive to demand. In bad times, employers will fail to take account of the poor opportunities available in the outside market to laid-off workers. In good times, workers will fail to take account of their own higher productivity in their current jobs when they contemplate quitting. Regrettable layoffs when demand is weak and regrettable quits when demand is strong are the outcome of practical limitations on contracts.

We do not consider this feature of the labor market a failure of free markets. Instead, it is an inevitable consequence of imperfect information. Our investigation suggests that some institutions that may seem arbitrary and even harmful—such as the involuntary layoff—are actually ways the economy achieves a second-best solution to the complex problem of allocating labor.
References


