Implicit Contracts, Labor Mobility, and Unemployment

By Richard J. Arnott, Arthur J. Hosios, and Joseph E. Stiglitz*

When workers' search efforts are unobservable, the provision of insurance against firm-specific shocks adversely affects their incentives to find better jobs. In consequence, the equilibrium contract prescribes low wages and underemployment to encourage workers to leave low-productivity firms; and it employs both quits and layoffs to induce separations, with the mix depending both on the relative efficiency of on- and off-the-job search and on the search-incentive effects of layoffs.

Wage rigidities have long played a central role in Keynesian explanations of unemployment. Though a major aim of implicit contract theory was to provide an explanation for these rigidities (Costas Azariadis, 1975; Martin Baily, 1974), it has become increasingly apparent that these rigidities do not comprise a complete explanation of unemployment; that is, while the simpler versions of implicit contract theory successfully explain real wage rigidities as a consequence of risk-averse workers' demand for insurance against variations in the value of their marginal product (VMP), they cannot explain unemployment.¹

To explain unemployment, one must explain, first, why reductions in demand induce layoffs rather than just reduced hours, and second, why those on layoff do not immediately secure employment elsewhere. Traditional implicit contract theory, by contrast, ignored both issues: it assumed that reductions in labor demand take the form of layoffs; and it assumed that all separated workers are immobile, due, say, to prohibitive mobility costs. The latter immobility assumption is particularly important since it implies that (i) laid-off workers are necessarily unemployed, (ii) quits are absent, (iii) high-demand firms cannot hire newly separated workers, and (iv) when optimally determined severance pay is provided, laid-off workers are better-off than retained workers. In fact, variable hours, layoffs, quits, and interfirn mobility are all prominent features of modern labor markets, and the prediction that workers prefer to be laid off is generally accepted as counterfactual.²

We propose here a model of implicit contracts which resolves most of these difficulties by relaxing the assumption that workers are immobile. We follow traditional contract theory in assuming that workers are risk-averse and firms are risk-neutral. As a result,

¹There are now a large number of surveys and critical reviews in this area, including George Akerlof and Hajime Miyazaki, 1980; Azariadis, 1979, Azariadis and Joseph Stiglitz, 1983; Oliver Hart, 1983; and most recently Sherwin Rosen, 1985.

²Though more recent developments in implicit contract theory, particularly those based on asymmetries of information (Azariadis, 1983; Sanford Grossman and Hart, 1981) have recognized the failure of the traditional versions to explain unemployment, they have failed to address the central problems raised above and have faced some further difficulties as well. For example, not only do these new theories obtain unemployment only under restrictive conditions but, when they do, they obtain it for all states of nature except the very best.
firms provide individuals with insurance. If labor mobility were costless and instantaneous, no insurance would be required against relative shocks (i.e., those which change the relative VMP's of labor in different uses) since each worker would move immediately to the job where his VMP is highest.

In fact, search is not only an individually costly activity, but is also a private activity that is costly to monitor. It is difficult for an employer to monitor his workers' search efforts and outside wage offers or even to verify whether his laid-off workers have become reemployed elsewhere. In turn, because the employer is unable to observe these search inputs and outcomes, the following moral hazard problems arise.

The first problem results when workers' job offers are private information, even when their search effort is observable. In this case, for each VMP realization, the firm can provide only a fixed severance payment to workers who quit or are laid off, and a given wage to workers who are retained. Consequently, the more insurance a firm provides against its own low-VMP draws (i.e., the greater the wedge between its workers' wage and VMP), the more likely it is that workers will refuse outside wage offers that exceed their current VMP but are less than their current wage, and hence the less efficient will be quit behavior.

The second moral hazard problem stems from the unobservability of search effort or intensity. In this case, the more insurance a firm provides against low-VMP draws, the less incentive a worker has to search for alternative employment and hence, once again, the less efficient will be quit behavior. Since it is privately (and socially) costly for individuals to remain where the value of their marginal product is low, the equilibrium insurance contract will in both situations provide incomplete insurance.

With private job-search information, and as a consequence of these moral hazard problems, the equilibrium implicit contract will be shown to prescribe relatively low wages and underemployment to encourage quits in bad times, and relatively high wages and overemployment to discourage quits in good times. In this setting, moreover, quits (induced by relatively low wages) and layoffs are alternative means of creating separations from low-VMP firms.

On one hand, reduced insurance (lower wages for retained workers) has the potential advantage of encouraging quits by the lowest search cost workers, effectively discriminating among otherwise indistinguishable workers. On the other hand, layoffs have the potential twofold advantage of forcing workers to use a different and likely more efficient off-the-job search technology and, when quits precede layoffs, of inducing more intensive on-the-job search by decreasing the returns to those who do not quit. As these two instruments, reduced wage insurance and layoffs, are imperfect substitutes, we expect to observe both quits and layoffs at individual firms.

In a limiting (and we would argue unrealistic) version of our model, in which there is no search, we obtain the standard counterfactual results. We therefore contend that an analysis of the role of implicit contracts in understanding unemployment must focus on questions of labor mobility and private job-search information. Regrettably, the simplest model that remedies the difficulties we have noted in the traditional framework must be somewhat complex: it must incorporate endogenous search; it must allow for the possibility of severance pay; and it must...
provide firms with a choice between layoffs and wage policy as means of encouraging the movement of workers from low- to high-VMP firms.

We begin in Section I with an analysis of firms’ wages, severance pay, and internal distortions when workers’ search efforts are private information and lay-off rates are given; Section I contrasts our results with earlier papers that assume either that separated workers are immobile or that job-search information is public. Section II presents a general model of layoffs and quits and compares implicit contracts that prescribe layoffs before and after on-the-job search; Section II also describes the equilibrium layoff rate in special cases with and without quits. Section III presents concluding remarks.

I. The Model

We develop a two-period model of a competitive economy which is buffeted by firm-specific shocks. During the ex ante period when there is still uncertainty, workers are free to join the firm whose employment contract offers the highest level of expected utility. At the beginning of the ex post period, after all random variables are realized, a fraction of each firm’s initial labor pool may be laid off as prescribed by the contract. Subsequently, after search, some retained workers and some laid-off workers secure new employment elsewhere. Finally, production takes place at the end of the ex post period. See Figure 1.

While this formulation may initially appear restrictive, it will become clear as we proceed that this is not the only possible interpretation of our model (see fn's 10 and 11), and that most of our qualitative results, for instance concerning incomplete insurance, underemployment, and the presence of layoffs and quits together, would hold under alternative sequencings as well. In particular, in Section II we show that our results are not substantially altered when on-the-job search precedes layoffs. Until then, however, the formulation in Figure 1 will be used to simplify the exposition.

The critical informational assumption of the model is that each worker’s ex post search effort and consequent job offers (if any) are privately known only by that worker.5

Firms are ex ante identical and risk-neutral, and labor is the only variable input to production. The VMP of a worker who supplies h units of labor is θh, where θ is the realization of a firm-specific shock; there are two equiprobable values of θ, θl, and θH, with θL < θH, where L(H) indexes the low-high productivity state. Workers have identical tastes and technical ability. As such the traditional literature, they are assumed to have no consumption good endowment and savings are not allowed.

Firms offer employment contracts to attract employees during the ex ante period which are designed to maximize expected profits, and workers choose among these contracts to maximize expected utility. A typical contract will be denoted by C = (wL, hL, r, s, (wH, hH)). Under this contract employed workers are paid wL for hL units of labor when θ = θL; a worker is retained with probability r in the low-VMP state and hence laid off with probability 1 – r; all workers are retained in the high-VMP state; and each laid-off worker is given the fixed nonnegative severance payment s.6

5 We also assume that each firm’s technology is observable by the firm and its workers but not by others. This gives risk-neutral firms on informational advantage over external agents, so that job-related insurance is provided to workers by employers rather than insurance companies.

6 Observe that asymmetric information has ruled out compensation conditional upon a worker’s search effort or outside job offer.

We require severance pay to be nonnegative because of the enforcement (default) problems associated with negative severance pay; the implications of negative severance pay are, however, described later in Section II. Severance pay for workers who quit is prohibited here only to simplify the discussion; see Section I, Part D, and fn. 23.

This list of contract parameters is not exhaustive. For example, we could allow firms to reemploy laid-off workers who are unsuccessful in finding alternative employment. In this case, if layoffs are to have the desired off-the-job search incentive effects, firms will not reemploy workers at terms that make unsuccessful search a desirable outcome. Since the same moral hazard issues emphasized throughout this paper would arise here as well, our results do not depend substantially on such
PERIOD 1

Workers choose among contracts specifying:
\[ h_l = \text{hours to be worked in the low-productivity state}, \]
\[ h_h = \text{hours to be worked in the high-productivity state}, \]
\[ w_l = \text{payment to employed workers in the low-productivity state}, \]
\[ w_h = \text{payment to employed workers in the high-productivity state}, \]
\[ r = \text{probability of being retained in the low-productivity state}, \]
\[ s = \text{severance payment to laid-off workers} \]

PERIOD 2

Realization of labor productivity:
\[ \theta \in \{\theta_l, \theta_h\}, \theta_l < \theta_h. \]

Firms make their employment decisions:
- if \( \theta = \theta_h \), worker is retained with certainty,
- if \( \theta = \theta_l \), worker is retained with probability \( r \) and laid off with probability \( 1-r \)

Workers make their search decisions:
- retained workers choose \( e(w, h) \) when \( \theta = \theta_l \);
- laid-off workers choose \( e(s) \)

Searching workers receive job offers:
- retained workers who accept their best offer quit to work elsewhere,
- reject their best offer remain employed at their initial firm,
- laid-off workers who accept their best offer become re-employed elsewhere,
- reject their best offer, or who receive no offers, remain unemployed

Production takes place.

FIGURE 1

The terms of employment offered by firms in the \textit{ex ante} period will depend on what alternative opportunities are available to workers in the \textit{ex post} period. We assume that there is a spot labor market characterized by a wage distribution, but in which all jobs require the same number of hours \( h > 0 \). Workers are concerned with the highest wage they can find, and this in turn depends on their search intensity and whether they are employed while searching.

We let \( F(z, e) \) and \( G(z, e) \) denote the distribution functions for the highest wage offers, \( z \), given search intensity \( e \), for those workers who are employed and unemployed, respectively; \( f \) and \( g \) are the corresponding density functions. \( F \) and \( G \) are assumed to satisfy: \( F(0, 0) = G(0, 0) = 1 \); and \( F_e, G_e < 0 \) for \( z > 0 \). In words, a worker cannot effortlessly receive an offer of employment; and the probability that his best offer will exceed \( z \) is an increasing function of his search intensity.

Taking their employment status as given, workers choose the level of search effort to maximize their expected utility. After substi-
tution, the resulting maximum expected utility enjoyed by retained workers is simply a function of their wages and hours, \( U(w, h) \); similarly, the maximum expected utility attained by laid-off workers is a function only of their severance pay, \( V(s) \). Thus, dropping the common multiplicative factor \( 1/2 \), total expected utility from contract \( C \) is:

\[
W(C) = rU(w_L, h_L) + (1-r)V(s) + U(w_H, h_H).
\]

Taking account of workers' search and quit behavior, firms' profits per retained employee are a function of the state \( \theta \) and of the worker's wages and hours, \( \pi(w, h, \theta) \), whereas the cost per laid-off worker is simply the severance payment, \( s \). Thus, total expected profit per employee is given by

\[
Z(C) = r\pi(w_L, h_L, \theta_L) - (1-r)s + \pi(w_H, h_H, \theta_H).
\]

As workers are freely mobile in the ex ante market, competition among firms will drive expected profits to zero. Hence the equilibrium contract is found by solving

\[
(P) \quad \max W(C) \text{ s.t. } Z(C) = 0.
\]

The remainder of this section describes and interprets the first-order conditions of this problem. Among the questions we ask are: Under what conditions will the wage in the bad state be below that in the good state, that is, when will there be at least partial wage flexibility? When will those who stay with the firm be underemployed? And finally, if there are layoffs, when will laid-off workers be worse-off than retained workers? Until Section II, the retention probability, \( r \), is taken as given.

\textbf{A. Derivation of Expressions for Expected Utility}

To answer these questions, we must first derive explicit expressions for a worker's expected utility when he is retained and when he is laid off. A worker's instantaneous utility function is given by \( \alpha(w, h) \). When a worker is retained, his expected utility, \( u(w, h, e) \), is the utility from his current job, \( \alpha(w, h) \), times the probability he stays with his current employer (i.e., does not get a better offer), plus the expected utility if he quits, \( \sigma(x, e) \), times the probability of quitting, minus the disutility, \( \beta(e, h) \), of searching for a better job at intensity \( e \):

\[
u(w, h, e) = \alpha(w, h)F(x, e) + \sigma(x, e)
\times (1 - F(x, e)) - \beta(e, h),
\]

where \( x = x(w, h) \) is the wage offer which makes the worker indifferent between quitting and staying, and is defined implicitly by \( \alpha(w, h) = \alpha(x, h) \). \( F(x, e) \) is the probability that a worker's highest wage offer is less than \( x \) given on-the-job search intensity \( e \), and the expected utility of a worker who quits is

\[
\sigma(x, e) = \left( \int_x^\infty \alpha(z, h) f(z, e) \, dz \right) / (1 - F(x, e)).
\]

We assume that \( \alpha \) is a strictly concave function satisfying \( \alpha_w > 0 \) and \( \alpha_h < 0 \) and that \( \beta \) is an increasing convex function of \( e \) satisfying \( \beta_e \geq 0 \) and \( \beta(0, h) = 0 \).

\textbf{Notes:}

\[8\] Workers are assumed not to quit to search, that is, \( U(w, h, e) \geq V(0) \). This condition is assumed to be satisfied, and will be ignored in the subsequent discussion.

\[9\] The additively separable structure and the imposition of \( \beta_e = 0 \) simplify the analysis but are otherwise inessential.

\[10\] The assumption, \( \beta_e \geq 0 \), is that an increase in hours worked cannot decrease the disutility of search effort. More than one interpretation of the model is possible and each involves a slightly different rationale for this assumption.

\[11\] Situations where an increase in hours worked leaves the worker with less time for search (or other nonwork) activities, obviously come to mind. One such interpreta-
pay for retained workers who quit is considered later in Section I, Part D.

Analogously, the expected utility of a laid-off worker who receives severance pay $s$ and expends $e$ on search equals

$$v(s, e) = \alpha(s, 0)G(y, e) + \int_0^\infty \alpha(z + s, h)g(z, e)\,dz - \beta(e, 0),$$

where $y = y(s)$ is the lowest acceptable wage offer and is defined by $\alpha(s, 0) = \alpha(y + s, h)$.

In the absence of search, observe that $u(w, h, 0) = \alpha(w, h)$ and $v(s, 0) = \alpha(s, 0)$.

Retained and laid-off workers' optimal on- and off-the-job search intensities are described, respectively, by $e(w, h) = \argmax u(w, h, e)$ and $e(s) = \argmax v(s, e)$. Hence their corresponding equilibrium utility levels are $U(w, h) = u(w, h, e(w, h))$ and $V(s) = v(s, e(s))$; and firms' equilibrium quit and profit functions per retained worker are, respectively,

$$q(w, h) = 1 - F(x(w, h), e(w, h)), \quad \pi(w, h, \theta) = (1 - q(w, h))(\theta h - w).$$

The quit rate is a function only of the wages and hours offered on the job and equals the probability of funding a higher utility job given the optimally chosen search effort, $e(w, h)$. Similarly, the job-finding rate for laid-off workers, $1 - G(y(s), e(s))$, is a function only of the severance payment $s$.

B. Wages and Production Efficiency

In this section, we shall show how firms' attempts to encourage mobility in the low-VMP state leads to incomplete insurance and, under normal conditions, to underemployment in this state.

From the first-order conditions of problem (P), with respect to wages and hours, we obtain

\begin{align*}
(1a) & \quad \frac{U_w(w_L, h_L)}{U_w(w_H, h_H)} = \frac{\pi_w(w_L, h_L, \theta_L)}{\pi_w(w_H, h_H, \theta_H)}, \\
(1b) & \quad \frac{U_h(w_i, h_i)}{U_w(w_i, h_i)} = \frac{\pi_h(w_i, h_i, \theta_i)}{\pi_w(w_i, h_i, \theta_i)},
\end{align*}

where $\theta_L > \theta_H > 0$.

The solution to (P) thus equates agents' marginal rates of substitution between wages across states, and between wages and hours within each state. Observe that these expressions measure the direct plus incentives-related substitution possibilities among the terms of employment for workers and firms;
and that the latter indirect effects take into account the adjustment of workers' search efforts and quit rates.

Returning to the basic equations defining $U$ and $\pi$, we can derive, for $i = L, H$,

\begin{align*}
(2a) \quad U_w(w_i, h_i) &= (1 - q(w_i, h_i)) \alpha_w(w_i, h_i), \\
(2b) \quad U_h(w_i, h_i) &= (1 - q(w_i, h_i)) \alpha_h(w_i, h_i) \\
&\quad - \beta_h(e(w_i, h_i), h_i), \\
(2c) \quad \pi_w(w_i, h_i, \theta_i) &= -(1 - q(w_i, h_i)) \\
&\quad - A_i q_w(w_i, h_i), \\
(2d) \quad \pi_h(w_i, h_i, \theta_i) &= (1 - q(w_i, h_i)) \theta_i \\
&\quad - A_i q_h(w_i, h_i),
\end{align*}

where

\[ A_i = \theta_i h_i - w_i. \]

Notice that $A_H$ and $-A_L$ are respectively equal to workers' implicit insurance premium and indemnity.

**Immobile Labor.**

In this case our model generates the standard results of full insurance and production efficiency; that is, substituting $q(w_i, h_i) = 0$ into (2), (1a, b) respectively simplify to

\[ \alpha_w(w_L, h_L) = \alpha_w(w_H, h_H), \]

\[ \frac{-U_h(w_i, h_i)}{U_w(w_i, h_i)} = -\frac{\alpha_h(w_i, h_i)}{\alpha_w(w_i, h_i)} = \theta_i, \]

\[ i = L, H. \]

As expected, the equilibrium contract equates workers' marginal utility of income across states and equates their marginal rate of substitution and marginal product in each state. Also, if either $\alpha$ is additively separable or the supply of labor is inelastic ($h_L = h_H$), we then get the well-known rigid contract wage result, $w_L = w_H$.

**Mobile Labor: Incomplete Insurance.** Suppose workers' equilibrium quit rates are nonzero in both states. In this case, a higher contract wage reduces the expected return to on-the-job search and, as a result, retained workers' search effort and consequent quit rate fall when the wage is raised, that is, $q_w(w_i, h_i) < 0$. Thus, making use of the fact that $A_L < 0 < A_H$, (1a) and (2a,c) give

\[(3) \quad \alpha_w(w_L, h_L) > \alpha_w(w_H, h_H). \]

That is, the marginal utility of income is higher in the low-VMP state than in the high-VMP state and, therefore, the optimal contract no longer provides complete insurance.

Since workers who have searched unsuccessfully are indistinguishable from those who make no attempt to search at all, firms will in part use the contract wage as an incentive device. Thus, when firms are subsidizing retained workers in the low-VMP state as part of the insurance package, $w_i$ is lowered (relative to the complete insurance situation) to encourage them to quit; and when workers are being taxed in the high-VMP state, $w_H$ is raised to discourage quits. In particular, if the utility function $\alpha$ is separable, or labor supply is inelastic, (3) gives $w_H > w_L$. 

\[ \frac{1}{\theta_1} = 1 - F(x, e) \] gives $q_x = -(F_x x + F_e e)$, where $F_x > 0$ and $F_e > 0$. From the definition of $x(w, h)$, $x_c = \sigma_c(w, h)/\sigma_c(x, h) > 0$; and from the f.o.c. $u_x(w, h, e) = 0$, where

\[ u_x = \alpha(w, h) F_x(x, e) + \int_x^\infty \alpha(z, h) f_z(z, e) \, dz \\
- \beta_e(e, h) \]

and $x = x(w, h)$, we have $q_x = -\alpha_x w F_x / u_x$. Then $q_w = -\alpha_w B$ where $B = (F_x / \alpha_x) - F_x / u_x > 0$ when the second-order condition, $u_{xx} < 0$, is satisfied.

\[ \frac{1}{\theta_1} \]

Previous work has recognized that wages will likely be raised to discourage quits when profits are high or turnover is costly (for example, Steve Salop, 1979), but has failed to make the symmetrical argument when profits are low. Separability is sufficient but not necessary for a positive correlation between real wages and VMP's.
Mobile Labor: Underemployment. We now examine whether our model generates underemployment. In the present search context, however, it is not immediately clear how this term should be defined. In particular, is there underemployment when \(-U_h/U_w < \theta\), that is, when the marginal rate of substitution between wages and hours is less than the marginal rate of transformation, taking into account induced search; or is there underemployment when \(-\alpha_h/\alpha_w < \theta\), that is, when the MRS between wages and hours along the instantaneous utility function is less than the MRT? Since it is easier to establish results concerning the relative magnitudes of \(-U_h/U_w\) and \(\theta\), we henceforth use the terms under and overemployment to describe situations, where \(-U_h/U_w < \theta\) and \(\theta > 0\), respectively, and say production efficiency obtains when the two are equal.\(^{15}\)

When quit rates are nonzero, (1b) and (2c,d) give

\[
\begin{align*}
\frac{U_h(w_i, h_i)}{U_w(w_i, h_i)} &= \frac{\theta_i - A_i q_h(w_i, h_i) / (1 - q(w_i, h_i))}{1 + A_i q_w(w_i, h_i) / (1 - q(w_i, h_i))}. \\
\end{align*}
\]

Earlier we noted that increasing wages reduces quit rates. If increasing hours also reduce quit rates, so that \(q_h < 0\), we get

\[
\begin{align*}
\frac{U_h(w_i, h_i)}{U_w(w_i, h_i)} &\geq \theta_i \text{ as } A_i \geq 0. \\
\end{align*}
\]

Therefore, when increasing hours worked reduces quits, the equilibrium contract prescribes underemployment (to encourage quits) in those states when workers are being subsidized, and overemployment (to discourage quits) when profits are strictly positive.\(^{16,17}\)

The property \(q_h < 0\) is sufficient but not necessary for (5). A necessary and sufficient condition for (5) is that the compensated quit derivative,

\[
\frac{dq}{dh} \bigg|_U = q_h + q_w \frac{dw}{dh} \bigg|_U,
\]

is negative; that is, there will be underemployment (overemployment) in those states where profits are negative (positive) if and only if workers’ compensated quit derivative with respect to hours is negative.

To see this, we consider a perturbation to \(\{w_i, h_i\}\) which keeps workers’ expected utility in state \(i\) fixed and calculate the effect on firm profits. At the optimum

\[
\begin{align*}
\frac{d\pi}{dh} \bigg|_U &= (1 - q) \left[ \theta - \frac{dw}{dh} \bigg|_U \right] - (\theta h - w) \frac{dq}{dh} \bigg|_U = 0. \\
\end{align*}
\]

Recalling

\[
\frac{dw}{dh} \bigg|_U = -U_h/U_w,
\]

and substituting, we can obtain

\[
\begin{align*}
\frac{-U_h(w_i, h_i)}{U_w(w_i, h_i)} &= \theta_i - A_i \frac{dq}{dh} \bigg|_U, \\
\end{align*}
\]

yielding the stated result.

\(^{15}\)From (a,b), \(U_h/U_w = \alpha_h/\alpha_w - \beta_h / [(1 - q) \alpha_w]\), and hence the two possible definitions of underemployment coincide if \(\beta_h = 0\), that is, if increasing hours worked does not affect the cost of searching (at any level of search effort). Normally, however, \(\beta_h > 0\) (see fn. 10) in which case \(-U_h/U_w > -\alpha_h/\alpha_w\). Accordingly, when there is underemployment under the first criterion (taking into account induced search), there is underemployment under the second criterion; and when there is overemployment under the second criterion (i.e., \(-\alpha_h/\alpha_w > \theta\)), there is overemployment under the first criterion.

\(^{16}\)Analogous results concerning the relative magnitudes of \(\alpha_h/\alpha_w\) and \(\theta\) follow straightforwardly with some obvious caveats (see fn. 15) and are omitted.\(^{17}\) These results are plausible yet appear to require \(q_h < 0\). It is possible, however, that \(q_h\) can be positive, even for well-behaved utility functions. There are two opposing effects at work (see the expression for \(q_h\) in fn. 18). On one hand, an increase in hours worked (at fixed wages) makes the job less attractive, and this increases quit rates. On the other hand, if search is time-intensive, then a worker who works more has less time to search, and this reduces quit rates.
In turn, it can be shown that

\[
\frac{dq}{dh} \bigg|_U = -\frac{B\beta_h}{(1-q)} - \frac{F_c\beta_{eh}}{u_{ee}},
\]

where \( B > 0 \) (see fn. 13), \( F_c < 0 \) and \( u_{ee} < 0 \). Therefore, a sufficient condition in our model for the compensated quit derivative to be negative is that both \( \beta_h \) and \( \beta_{eh} \) be positive, that is, that increasing hours worked increases both the total and marginal costs of search. Observe that the compensated quit derivative is negative under less restrictive conditions than are required for \( q_h < 0 \).\(^{18}\) Observe also that the equilibrium contract satisfies \( -U_h/U_w = \theta \) and hence exhibits production efficiency whenever the compensated quit derivative is zero.

The following properties of these results are noteworthy. First, the occurrence of under- versus overemployment in our model does not depend on whether firms are risk-averse or, with reference to either \( u(w, h, e(w, h)) \) or \( \alpha(w, h) \), whether leisure is a normal good, or finally, whether \( \alpha(w, h) \) has a specific functional form, such as \( \alpha = \mu(w) - \gamma(h) \) or \( \alpha = \mu(w - \gamma(h)) \).\(^{20}\) Whether there is under- or overemployment depends

\[^{18}\text{It can be shown that } q_h = -\alpha_h B + (F_c\beta_{eh} + u_{ee}), \text{ where, as in fn. 13, } B = -q_h/\alpha_w. \text{ Also, from (2a,b),}\]

\[
-U_h/U_w = -\frac{\alpha_h}{\alpha_w} + \frac{\beta_h}{\alpha_w(1-q)} = \frac{dw}{dh} \bigg|_U,
\]

and therefore, substituting \( q_w = -\alpha_w B \) and the above expressions for \( q_h \) and \( -U_h/U_w \) into

\[
\frac{dq}{dh} \bigg|_U = q_h + q_w \frac{dw}{dh} \bigg|_U
\]

gives (7).

\[^{19}\text{Since } q_h = -\alpha_h B - F_c\beta_{eh}/u_{ee}, \text{ where } \alpha_h < 0, (7) \text{ is negative under less restrictive conditions than are required for } q_h < 0. \]

\[^{20}\text{See Russell Cooper, 1983; Hart, 1983; and Stiglitz, 1986, for a description of the critical role of preferences in generating production inefficiencies in models where the firm, but not the worker, can observe the state of nature.}

Instead only upon certain weak properties of the utility function (for example, \( \beta_{eh} > 0 \)) and upon the contractual availability of insurance for workers (\( A \neq 0 \)).

Second, (5)–(7) indicate that under reasonable conditions underemployment will occur only in states where workers are being subsidized, with the amount of underemployment positively related to the size of the subsidy; and that overemployment will be reserved for profitable states, with the amount of overemployment again responding to the level of profits. Thus, as seems reasonable, underemployment occurs in bad times and overemployment in good times.\(^{21}\)

C. On Severance Pay

Severance pay is oftentimes portrayed as the Achilles’ heel of implicit contract theory. The original models largely ignored severance pay. Later work introduced severance pay but obtained the counterfactual result that workers prefer to be laid off rather than retained; that is, contracts which equate (expected) marginal utilities of income across all states result in higher (expected) utilities among those on layoff, in which case one would expect contracts to contain reverse seniority layoff clauses. In this section we show, on the contrary, that under reasonable conditions concerning labor mobility and preferences, laid-off workers are worse off than retained workers. Before presenting our results, we review the standard ones.

The first-order conditions of problem (P) yield

\[
(8) \quad -U_w(w_i, h_i)/\pi_w(w_i, h_i, \theta_i) = V_s(s)
\]

whenever the optimal severance payment \( s \) is strictly positive. (Later, we identify cases where \( s = 0 \).) Observe that when both on- and off-the-job search occur, (1), (2), and (8)

\[^{21}\text{By contrast, earlier models generated either underemployment in all high-VMP states except the highest (Azariadis, 1983; Grossman and Hart, 1981) or overemployment in all low-VMP states except the lowest (V. V. Chari, 1983; Cooper, 1983; Jerry Green and Charles Kahn, 1983).} \]
give
\[ \alpha_w(w_L, h_L) > V'_s(s) > \alpha_w(w_H, h_H) \]
for \( A_L < 0 < A_H \),
which is the obvious extension of our earlier partial insurance result.

**Reverse Seniority.** There are two versions of the reverse seniority result. First, suppose employed and laid-off workers are both immobile (do not search) so that (8) becomes
\[ \alpha_w(w_L, h_L) = \alpha_s(s, 0). \tag{9} \]
The optimal (insurance) contract equates the marginal utilities of income of retained and laid-off workers. Differentiating \( \alpha_w = \) constant, gives \( dw/dh = -\alpha_{wh}/\alpha_{ww} \), and hence \( d\alpha/dh = \alpha_h = \alpha_w \alpha_{wh}/\alpha_{ww} \) for fixed \( \alpha_w \). Therefore, (9) implies that
\[ \alpha(w_L, h_L) < (>) \alpha(s, 0), \]
when leisure is a normal (inferior) good; and hence **immobile workers prefer to be laid off when leisure is normal** (Cooper, 1983; Stiglitz, 1986).

Second, suppose employed workers are again immobile, and that laid-off workers are now mobile but face uncertain job opportunities. To simplify, suppose the utility function is separable, \( \alpha(w, h) = \mu(w) - \gamma(h) \). In this case, (8) again yields the result that retained and laid-off workers’ (expected) marginal utilities of income are equal, that is,
\[ \mu'(w_L) = \mu'(s) G(y, e) \]
\[ + \int_y^\infty \mu'(s + z) g(z, e) \, dz, \]
where \( e = e(s) \). Equation (10) implies that \( \mu(w_L) \) will be greater than (equal to, less than)
\[ \mu(s) G(y, e) + \int_y^\infty \mu(s + z) g(z, e) \, dz, \]
as \( \mu(\ ) \) is an increasing (constant, decreasing) absolute-risk-aversion utility function.\(^2\)

Thus, whenever search entails zero disutility and hours worked is not a variable, workers with decreasing absolute risk-aversion utility functions prefer layoffs with severance pay.

**Costly Search.** With costly search, however, the presumption that laid-off workers are better-off no longer obtains. There are two reasons for this. First, if search costs are not “completely pecuniary” (i.e., do not enter the utility function only additively with income), then providing full-income insurance (by equating expected marginal utilities) will not fully compensate for the costs and risks of search. Second, retained workers in low-VMP firms will, we have argued, typically have their hours reduced to encourage them to search more; as a result, hours may be reduced below the “standard” hours at alternative employment so that, even with full-income insurance, retained workers’ expected utility may be higher than that of laid-off workers.

To see these results, first suppose every laid-off worker can effortlessly (with \( e = 0 \)) secure a new job which pays \( z \) for \( h \); since there are no incentive problems here, the optimal severance payment will satisfy
\[ \alpha_w(w_L, h_L) = \alpha_s(s + z, h^\prime). \tag{11a} \]
It then follows that laid-off workers will be worse-off when leisure is normal whenever the spot job entails more work (\( h > h_L \). Now, suppose instead that this alternative job can be secured only by exerting some fixed effort \( \hat{e} > 0 \); assuming that laid-off workers choose to search, retained workers are better-off when
\[ \alpha(w_L, h_L) > \alpha(s + z, h^\prime) \]
\[ + \beta(\hat{e}, 0). \tag{11b} \]

\(^{22}\)Whether the expected utility of income is greater or less for laid-off than for retained workers depends upon whether utility is a concave or convex function of marginal utility. Recognizing that \( s + z \) is a mean marginal utility preserving spread of \( w_L \), this result follows as a corollary to Peter Diamond and Stiglitz, 1974. See also Haruo Imai, John Geanakoplos, and Ito Takatoshi, 1981.
Whenever search costs are in part nonpecuniary, it is clear that (11a) and (11b) can both be satisfied under a wide range of conditions.

Thus, in the special case where employed workers are immobile we have the following result: Provided \( i \) search costs are nonpecuniary (and accordingly, severance payments do not fully cover them), \( ii \) leisure is normal, and \( iii \) there are significant probabilities of being reemployed elsewhere at jobs requiring at least as much work as at their ex ante firms, laid-off workers will be worse-off than retained workers; and this will be true even in the presence of decreasing absolute risk aversion.

More generally, when the disutility of search is not an additively separable component of utility and retained workers also search on the job, it is clear that one can infer little from (8) about the relative magnitudes of \( U(w_l, h_L) \) and \( V(s) \).

D. Severance Pay for Retained Workers

Our results are not substantially altered if firms provide severance pay for retained workers who quit in bad states. Such severance pay would be used to encourage quits. However, so long as reducing hours worked reduces the cost of on-the-job search, underemployment will still occur in low-VMP states; so long as reducing wages increases the return to search, firms will not provide complete income insurance; so long as search costs are not completely additive with income, the severance pay will not fully compensate for effort expended on search; and finally, so long as there is imperfect information about job offers, the distortions we described earlier with respect to quit behavior will remain.\(^{23}\)

\(^{23}\)Suppose firms pay \( b \) to workers who quit. In this case, only those workers who receive an outside offer \( z \geq x - b \) will quit, where \( x = x(w, h) \) is again defined by \( a(w, h) = \alpha(x, h) \). As a result, \( u(w, h, e) \) is replaced by

\[
u(w, h, e, b) = a(w, h)F(x - b, e) + \sigma(x, e, b) \times (1 - F(x - b, e)) - \beta(e, h),\]

where

\[
\sigma(x, e, b) = \int_{x-b}^{a(z+b, h)} f(z, e) \, dz / (1 - F(x - b, e)).
\]

Hence the optimal search intensity and resulting quit rate are given by

\[
e(w, h, b) = \arg\max u(w, h, e, b),
\]

\[
q(w, h, b) = 1 - F(x - b, e(w, h, b)),
\]

where \( q_b > 0 \). Similarly, the resulting utility and profit functions are given by

\[
U(w, h, b) = u(w, h, e(w, h, b), b),
\]

\[
\pi(w, h, b) = (1 - q(w, h, b))(\theta h - w) - q(w, h, b).\]

As \( U_u \) and \( U_d \) are again described by (2a,b), and \( \pi_u \) and \( \pi_d \) by (2c,d), where \( A = \theta h - w + b \), our earlier results on wages and production inefficiencies go through exactly as before. See Ito, 1986; and Charles Kahn, 1985, for further results on quits and severance pay.

\(^{24}\)See Ken Chan and Yannis Ioannides, 1982; Mark Lowenstein, 1983; Rosen, 1985; Andrew Weiss, 1980; and Stiglitz, 1986.
the firm provides severance pay and whether layoffs precede or follow quits.

We begin by formulating a general model using the structure and notation of the preceding section; this model allows us to identify the principal determinants of the equilibrium layoff rate and to verify that whether layoffs precede or follow on-the-job search is not a critical factor. Section II, Part B particularizes the model to derive specific results.

A. A General Formulation

Suppose the ex post period is divided into two subperiods, of lengths $t$ and $1 - t$, such that workers can search and work in both subperiods in exactly the same manner as described earlier in Section I. Layoffs, meanwhile, can occur only once, at the beginning of either the first or second subperiod. We consider two cases; (all) layoffs precede quits ($LQ$) when layoffs occur at the beginning of the first subperiod (see Figure 2(a)) and (some) quits precede layoffs ($QL$) when layoffs occur at the beginning of the second subperiod (see Figure 2(b)). Section I considered the special case $LQ$ with $t = 1$.

Earlier we let $U = F \alpha + (1 - F) \sigma - \beta$ denote the expected utility of a retained worker where $1 - F$ is the quit rate that results with the optimal search effort. Now, using superscripts to denote subperiods ($i = 1, 2$) and subscripts to denote states ($j = L, H$), we adapt this notation and the corresponding expression for laid-off workers as follows. Let $tU^1_j$ and $(1 - t)U^2_j$, where $U^i_j = \alpha^i_jF^i_j + \sigma^i_j(1 - F^i_j) - \beta^i_j$, respectively, denote the expected utilities of a retained worker in state $j$ and in subperiods 1 and 2; and let $tV^1$ and $(1 - t)V^2$, defined analogously, respectively denote the expected utilities of a laid-off worker in subperiods 1 and 2.

When layoffs precede quits, total expected utility is

(12a) $W^{LQ} = r\hat{U}_L + (1 - r)\hat{V} + \hat{U}_H$,

where $\hat{U}_L$, $\hat{V}$, and $\hat{U}_H$ are the expected utilities of a worker conditional on his being retained in the low-VMP state, laid off in the low-VMP state, and retained in the high-VMP state, respectively, at the beginning of the first subperiod. Employing the notation introduced above, we have

$$\hat{U}_j = tU^1_j + (1 - t)(F^1_jU^2_j + (1 - F^1_j)\sigma^1_j)$$

$$\hat{V} = tV^1 + (1 - t)(G^1V^2 + (1 - G^1)\sigma^1)$$.

For example, the expected utility of a worker retained at a low-VMP firm equals his expected utility in the first subperiod, $tU^1_L$, plus the expected utility in the second subperiod conditional on having quit during the first, $(1 - t)\sigma^1_L$, times the probability of quitting during the first subperiod, $(1 - F^1_L)$, plus the expected utility in the second subperiod conditional on having not quit during the

---

**Figure 2**
first, \((1 - t) \Delta U_L^2\), times the probability of staying through the first subperiod, \(F_L^1\). Total expected profit per worker is

\[
Z^{LO} = r \hat{\sigma}_L - (1 - r) s + \hat{\sigma}_H, \tag{12b}
\]

where \(\hat{\sigma}_j = t \sigma_j^1 + (1 - t) F_j^1 \sigma_j^2\), and \(\sigma_j^r = F_j^1 (\theta_j h_j + w_j)\).

When the layoff decision is made at the beginning of the second subperiod, so that \textit{quits precede layoffs}, total expected utility is

\[
W^{QL} = t U_L^1 + (1 - t) \times \left[ F_L^1 [r U_L^2 + (1 - r) V^2] + (1 - F_L^1) \sigma_L^1 + \hat{U}_H \right]. \tag{13a}
\]

Observe that all workers at low-VMP firms are retained during the first subperiod and that from among those who stay to the second subperiod, a fraction \((1 - r)\) are laid off. Total expected profit per worker is

\[
Z^{QL} = t \sigma_L^1 + (1 - t) \times F_L^1 [r \sigma_L^2 - (1 - r) s] + \hat{\sigma}_H. \tag{13b}
\]

The optimal contract when layoffs precede quits is found by maximizing (12a) subject to the constraint that (12b) is zero; and the optimal contract when quits precede layoffs is found by maximizing (13a) subject to the constraint that (13b) is zero. Whether the equilibrium contract prescribes layoffs before or after on-the-job search is determined by comparing the solutions to these two problems. Fortunately, as shown below, it is not necessary to make this comparison to determine whether layoff rates are ever positive.

When layoffs are absent, so that the retention probability equals one, the payoffs for workers and firms are the same whether layoffs precede or follow quits; that is, at \(r = 1\), (12a) and (13a) are the same and (12b) and (13b) are the same. Therefore, if the optimal layoff rate is positive for one sequence of events and is zero for the other, workers must strictly prefer the sequence that delivers the positive layoff rate. In particular, therefore, the optimal contract when layoffs precede quits strictly dominates the optimal contract when quits precede layoffs whenever the former prescribes layoffs but the latter does not.

Although the determinants of layoffs when layoffs precede and follow quits are different, sufficient conditions for \(r < 1\) under either sequence are also sufficient when the sequence is itself a choice variable. As a result, we can investigate each case separately.

When layoffs precede quits, the derivative of the Lagrangian corresponding to (12), \(L^{LO} = W^{QL} + \lambda Z^{LO}\), with respect to the retention rate is

\[
\frac{\partial L^{LO}}{\partial r} = \hat{U}_L - \hat{V} + \lambda (\hat{\sigma}_L + s), \tag{14a}
\]

where \(\lambda\) is the multiplier corresponding to the zero-expected profit constraint and, from the other first-order conditions, is interpreted as the marginal utility of income. Retaining an extra worker has two effects: it makes that worker better off by \(\hat{U}_L - \hat{V}\); and it increases the firm's subsidization costs in the low-productivity state by \(- (\hat{\sigma}_L + s)\) which, since this cost is passed on, lowers the expected utility of all workers.

Thus, taking the other contract parameters as given, we expect more efficient \textit{on-the-job} search to increase the benefit of retaining an extra worker by raising the expected utility on the job, \(\hat{U}_L\), and to decrease the cost of retaining an extra worker by inducing more quits and hence lowering the expected subsidy, \(- \hat{\sigma}_L\). By contrast, we expect more efficient \textit{off-the-job} search to decrease the benefit of retaining an extra worker by raising the expected utility off the job, \(\hat{V}\). Therefore, as off-the-job search becomes relatively more efficient, we expect that the equilibrium mix of quits and layoffs will shift from mostly quits to mostly layoffs.\(^{22}\) Section II, Part B

\(^{22}\)This shift will be continuous and interior solutions \((0 < r < 1)\) will result if workers' total expected utility is strictly concave in the retention rate, \textit{mutatis mutandis}; this function will normally be strictly concave due to the strict concavity of the underlying instantaneous utility function, \(\alpha(w, h)\) (see Section II, Part B).
and the Appendix, (Part A) derive these results using a version of the above model that entails a specific formalization of the notion of “more efficient search.”

When quits precede layoffs, there are three factors at work determining the layoff rate: First, there is the insurance-mobility-cost tradeoff described above that is the sole determinant when layoffs precede quits. Second, there is the screening effect of on-the-job search which tends to decrease the layoff rate. When workers differ in their search ability, allowing on-the-job search to precede layoffs screens out the efficient searchers and leaves the least mobile workers behind. In effect, this initial screening makes subsequent off-the-job search less efficient, on average, which in turn makes subsequent layoffs less attractive. Finally, there is the incentive effect of layoffs which tends to increase the layoff rate. Allowing layoffs to follow on-the-job search will make workers initially search more intensively on the job because unsuccessful search exposes them to the subsequent risk of being laid off; thus, by inducing quits and thereby decreasing the number of subsidized workers at low-VMP firms, the benefits of layoffs should be enhanced.

These incentive and screening effects are unique to contracts which prescribe on-the-job search before layoffs because, in this situation, search effort during the first subperiod will depend on the retention rate during the second (see Figure 2(b)). The derivative of the Lagrangian corresponding to (13), \( L^{QL} = W^{QL} + \lambda Z^{QL} \), with respect to the retention rate is

\[
\frac{\partial L^{QL}}{\partial r} = (1 - t) F_1^{\frac{2}{2}} [U^{2} - V^{2} + \lambda (\pi^{2} L + s)] \\
+ (1 - t)(1 - r) F_1^{\left(\frac{2}{2}\right)} \\
+ \lambda \left( t \frac{\partial \pi^{1}}{\partial r} + (1 - t) \right) \\
\times \left[ r\pi^{2} L - (1 - r) s \right] \left( \frac{\partial F_1^{\frac{2}{2}}}{\partial r} \right).
\]

Conditional on not quitting during the first subperiod, the first term in (14b) represents the same net benefit described earlier in (14a). The second term in (14b) represents the screening effect and the third term represents the incentive effect. An example in the Appendix (Part B) details these effects and confirms their predicted signs.

**B. The Equilibrium Layoff Rate**

In this part and the Appendix we examine some special versions of our general formulation. In all instances, we assume that workers’ preferences and search technologies are characterized as follows:

**ASSUMPTION 1**: The search disutility function takes on the form \( \beta(e, h) = e \); with \( \beta_k = 0 \) we know, from (6) and (7), that long-term and spot employment contracts will exhibit production efficiency.

**ASSUMPTION 2**: The income-leisure utility function takes on the special form \( \alpha(w, h) = \mu(w - h^2/2) \), where \( \mu(0) = 0 \). We shall refer to \( w - h^2/2 \) as the “net income” of a job, that is, the wage income net of labor costs. Since production efficiency implies \( -\alpha_w / \alpha_h = \theta \), workers will supply \( h = \theta' \) whenever \( \theta = \theta' \). Thus, \( \alpha(w, h) = \mu(w - \theta^2/2) \) in equilibrium. Observe that the assumed form of the utility function implies \( h_H > h_L \) and hence results in work-sharing.

**ASSUMPTION 3**: We consider the following special case of the search process introduced earlier: An employed (laid-off) worker either expends at least \( e_{on} \) (\( e_{off} \)) to become fully informed of all employment opportunities, or foregoes search and remains uninformed (and immobile). The requisite effort levels, \( e_{on} \) and \( e_{off} \), are fixed for each worker but can vary across workers with the same employment status.

Assumption 3 allows us to endogenezie workers’ alternative job offers in the \textit{ex post} spot market as follows.

With firm-specific shocks and constant returns, there will always be a group of high-VMP firms that is willing to hire workers on
spot contracts. Since all searching workers have perfect information, the resulting ex post equilibrium must conform to the classical zero-profit competitive outcome; hence all spot contracts must pay $z = \theta_H^2$ for $h = \theta_H$ units of labor, so that $\alpha(z, h) = \mu(\theta_H^2/2)$. Thus, using our earlier notation, the on-the-job search technology corresponding to Assumption 3 can be formally described as follows:

$$F(z, e) = \begin{cases} 
1 & z \geq 0, \quad e < e_{on}, \\
1 & z \geq \theta_H^2, \quad e \geq e_{on}, \\
0 & z < \theta_H^2, \quad e \geq e_{on},
\end{cases}$$

where $F(z, e)$ is the probability that a worker’s best offer is $z$ or less, given search effort $e$ on the job. The off-the-job distribution function, $G(z, e)$, is described similarly with $e_{off}$ replacing $e_{on}$. Layoffs Without Quits.\textsuperscript{26} In this subsection we assume that layoffs precede quits and search occurs at most once ($t = 1$), that workers are identical (the search cost distributions are degenerate), and that $e_{on} \geq \mu(\theta_H^2/2)$ while $e_{off} = \hat{e} < \mu(\theta_H^2/2)$. In this benchmark case, it is clear that retained workers will choose to forego search (and not quit); the Appendix allows nondegenerate cost distributions, so that some retained workers quit, and compares situations where layoffs precede and follow quits.

If laid-off workers chose to forego search, so that $V(s) = \mu(s)$, it is as if they are immobile; and, when laid-off workers are immobile, it is easy to show that the optimal layoff rate is zero.\textsuperscript{27} Hence layoffs occur only when the level of severance pay is set so that laid-off workers search. Retained and laid-off workers’ expected utilities are therefore, respectively, $U(w_i, h_i) = \mu(w_i - \theta_i^2/2)$ ($i = L, H$) and $V(s) = \mu(s + \theta_H^2/2 - \hat{e})$. Since the equilibrium contract maximizes total expected utility,

$$W = r\mu(w_L - \theta_L^2/2) + (1 - r)$$

$$\times (\mu(s + \theta_H^2/2 - \hat{e}) + \mu(w_H - \theta_H^2/2),$$

subject to the zero-expected profit constraint,

$$Z = r(\theta_L^2 - w_L) - (1 - r)s + (\theta_H^2 - w_H) = 0,$$

we see that employed workers will receive full insurance, that is, $\mu(w_L - \theta_L^2/2) = \mu(w_H - \theta_H^2/2)$. If, as we assume here, negative severance payments are not permitted, it can be shown that the equilibrium value of $s$ must be zero.\textsuperscript{28} (Negative values are discussed below.)

The model’s remaining equilibrium properties are derived as follows: Solving $Z = s = 0$ and $w_L - \theta_L^2/2 = w_H - \theta_H^2/2$ gives

$$w_i - \theta_i^2/2 = (r\theta_i^2 + \theta_H^2)/(2(1 + r)) = \delta(r),$$

so that workers’ expected utility $W$ can be rewritten as

$$W(r, \hat{e}) = (1 + r)\mu(\delta(r))$$

$$+ (1 - r)\{\mu(\theta_H^2/2) - \hat{e}\}.$$  

Observe that $W_{rr} < 0$ for $0 \leq r \leq 1$. There are

\textsuperscript{26}A more detailed exposition of the material in this section, including some omitted proofs, is contained in Arnott, Hosios, and Stiglitz, 1983. The model in this section resembles Baily’s, 1977.

\textsuperscript{27}This is because a contract that pays $w$ for $h$ hours to retained workers, pays $s$ to immobile laid-off workers and uses the retainment rate $r < 1$ is dominated by the contract that pays $rw + (1 - r)s$ for $rh$ hours to retained workers and retains all workers, that is, $r(\theta_h - w) - (1 - r)s = \theta_h - (rw + (1 - r)s)$ while $ra(w, h) + (1 - r)a(s, 0) < a(rw + (1 - r)s, rh)$.

\textsuperscript{28}Since $\mu(\theta_H^2/2) - \hat{e} > \mu(0) = 0$, every laid-off worker will search when $s = 0$. To verify that $s = 0$, note that $w_l - \theta_L^2/2 = w_H - \theta_H^2/2$ and $\theta_H^2 > \theta_L^2$ imply $\theta_H^2 > w_H > \theta_L^2 - w_L$, so that $Z = 0$ and $s \geq 0$ give $\theta_H^2 - w_H \geq 0$. Therefore, in equilibrium, the profit per employee at high-VMP firms with this implicit contract is greater than or equal to zero, while the profit per employee at firms with spot contracts is zero. Since the spot contract maximizes $\mu(s + w - h^2/2)$ subject to $\theta_H h - w = 0$, and since the optimal implicit contract maximizes $\mu(w_H - h_H^2/2)$ subject to $\theta_H h_H - w_H = 0$, for some $\hat{e} \geq 0$, $s > 0$ implies $\mu(s + \theta_H^2/2) > \mu(w_H - \theta_H^2/2)$. Hence a nonpositive $s$ value is required to equate retained and laid-off workers’ marginal utilities of income, $\mu'(w_H - \theta_H^2/2)$ and $\mu'(s + \theta_H^2/2)$. 

Copyright © 2001 All Rights Reserved
Figure 3

two corner solutions to consider: if $W_r \leq 0$ at $r = 0$, no one is retained; and if $W_r \geq 0$ at $r = 1$, everyone is retained. To determine when these solutions occur, define the critical effort levels, $e_u(\theta_L^2)$ and $e_1(\theta_L^2)$ by the first-order conditions

$$W_r(1, e_u(\theta_L^2)) = 0 = W_r(0, e_1(\theta_L^2)).$$

These effort levels are represented in Figure 3. Thus, if $\hat{e} \geq e_u(\theta_L^2)$, $r = 1$; if $\hat{e} \leq e_1(\theta_L^2)$, $r = 0$; and if $e_1(\theta_L^2) < \hat{e} < e_u(\theta_L^2)$, $0 < r < 1$ and it can be shown that $dr/d\hat{e} > 0$ and $dr/d\theta_L > 0$.

The tradeoff between risk-sharing considerations and search costs is clear. Off-the-job search becomes less expensive and hence more efficient, or the negative productivity shock becomes larger (as $\theta_L$ falls), the balance tips toward layoffs and search, and away from work-sharing; any combination of $\hat{e}$ and $\theta_L^2$ below (above) the shaded area in Figure 3 involves layoffs (work-sharing) and no work-sharing (layoffs).

For intermediate values of $\hat{e}$, however, some workers are retained to work-share in the low-VMP state while the remainder are laid off with zero-severance pay and undertake search.

On Severance Pay (Again). With constant returns to labor and identical workers, one might have conjectured that the equilibrium employment contract would treat all workers identically ex post, and hence that the equilibrium layoff rate must be a corner solution, either zero or one. In fact, of course, interior solutions are possible, as demonstrated above.

Thus far we have restricted severance pay to the nonnegative (due to enforcement problems) and hence one might have conjectured further that the binding nonnegativity constraint in our analysis is the source of the interior solution. It can be shown that this is not in fact the case: When negative severance pay is allowed, there is full insurance and interior layoff solutions ($0 < r < 1$) are possible whenever search involves some nonpecu-

---

29 These corner solutions are familiar: when workers are freely mobile ($\hat{e} = 0$), one-period implicit contract models must replicate spot auctions; and when workers are immobile ($\hat{e} > \mu(\theta_D^2/2)$), the concavity of $\mu$ ensures that work-sharing strictly dominates layoffs.
niary and hence uninsurable costs. With complete insurance and only pecuniary search costs, however, the equilibrium layoff rate simply implements the standard "first-best" allocation ($r = 0$ or 1).

* Layoffs with Quits.* The model described above was structured so that there would be no on-the-job search (and hence no quits). This model allowed us to identify a simple inverse relationship between the cost of off-the-job search and the equilibrium layoff rate and, as a special example, confirmed that interior layoff solutions are a feature of the general class of models described earlier in Section II, Part A. Nevertheless, this model failed to capture the tradeoff between voluntary separations (quits induced by lower pay and underemployment) and involuntary separations (through layoffs).

In the Appendix we allow quits and describe the factors that determine the equilibrium mix of quits and layoffs. To capture the sorting effect of quits in the simplest way, we modify the above model by assuming that search costs differ among workers. In particular, we assume that immediately after learning his employment status, each retained worker draws a private value for $e_{on}$ from a distribution with c.d.f. $I(e_{on})$; and each laid-off worker draws a private value for $e_{off}$ from a distribution with c.d.f. $J(e_{off})$. (Recall, retained (laid-off) workers must expend $e \geq e_{on}$ ($e \geq e_{off}$) to find alternative employment.)

Although quits expose workers to less down-side risk and induce voluntary sorting when workers differ in the ability at search, we show that layoffs are doubly beneficial: layoffs force workers to search off the job, which is likely more efficient than on-the-job search; and, when on-the-job search precedes layoffs, they induce more intensive on-the-job search and hence more turnover at low-productivity firms. The former benefit is highlighted in Part A of the Appendix, by comparing the effects of different $I(\cdot)$ and $J(\cdot)$ distributions on the layoff rate, using a version of the model in which layoffs precede quits; the latter incentive effect of layoffs and the companion screening effect of on-the-job search are highlighted in Part B with a version of the same model in which quits precede layoffs. Since workers prefer the quit-layoff sequence whose optimal contract prescribes layoffs, however, the analyses in this section and in the Appendix indicate that whether layoffs precede or follow quits, one should nevertheless expect to observe both quits and layoffs at low-productivity firms.

**III. Concluding Remarks**

A complete theory of unemployment should answer the following questions: How do we explain the degree of observed wage flexibility? How do we explain the form of employment changes, for example, variable hours versus layoffs? How do we explain which workers get laid off? How do we explain the level of severance pay? Why does severance pay not fully compensate laid-off workers? Why do workers prefer to be retained rather than laid off? Why are some unattached workers unable or unwilling to secure employment elsewhere?

While existing implicit contract models answer some of these questions, they fail to answer others and, in some instances, yield counterfactual implications. We have argued that costly search coupled with firms' inability to monitor workers' search activities can provide insights into each of the questions listed above; they can explain partial wage insurance, underemployment at low-VMP firms, layoffs, and quits, slugging: interfirm mobility, and equilibrium unemployment.

A complete theory of unemployment should also indicate whether the allocation of resources is efficient. From the literature on equilibrium with incomplete markets and imperfect information, it should be evident
that although the equilibrium contracts described in this paper are “locally efficient” (i.e., given the actions of all other firms in the economy, these contracts maximize workers’ expected utility subject to a zero-profit constraint), the market equilibrium, and the corresponding “natural rate” are generally not constrained Pareto-efficient.\footnote{\(31\)}

To summarize: By highlighting the implications of private search information, this paper has resolved several of the outstanding conundrums in the implicit contract literature and has also provided a unified treatment of variable hours, layoffs, severance pay, quits, and unemployment. The reason why it is necessary to formulate a model which incorporates all of these features should be apparent; without doing so, one cannot be sure that whatever results one obtains are not due to artificial restrictions (for example, that there is no severance pay, no on-the-job search, or that layoffs precede search on the job).

Our model can be viewed as an extension of the standard theory to take into account that what is at risk is not being laid off, per se, but rather being laid off and not rehired. As an extension of this earlier theory, however, some common issues remain; how, for instance, are such contracts enforced? While these problems remain important, their resolution will not, we believe, alter the qualitative insights provided by our analysis.

APPENDIX

In this Appendix we examine different versions of the model described in Section II, Part B; to simplify the remaining analysis we ignore severance pay.

A. Layoffs Precede Quits. We assume that immediately after learning his employment status, each worker draws a private value for \(e_{on}\) if he is retained or \(e_{off}\) if he is laid off. We suppose that retained workers’ \(e_{on}\) values are drawn from a distribution with continuous c.d.f. \(I(e_{on})\) and support \([0, \mu(\theta_{H}/2)]\) and that unattached workers’ \(e_{off}\) values are drawn from another distribution with the same support and a possibly different mean value \(\bar{e}\).

Assuming once again that retained workers search to quit (rather than quit to search) and substituting the equilibrium labor supplies \(h_{s} = \theta_{s}\), agents’ expected utilities are again described by \(W(C)\) and \(Z(C)\) as in Section I, except that

\[
U(w_{s}, h_{s}) = (1 - I(x_{s}))\mu(w_{s} - \theta_{s}/2) + I(x_{s})\mu(\theta_{H}/2) - \int_{0}^{x} edl(e),
\]

\[
V(0) = \mu(\theta_{H}/2) - \bar{e},
\]

\[
\pi(w_{s}, h_{s}, \theta_{s}) = (1 - I(x_{s}))(\theta_{s} - w_{s}),
\]

where \(x_{s} = \mu(\theta_{H}/2) - \mu(w_{s} - \theta_{s}/2) > 0\). \(x_{s}\) is the difference in utility from employment in a new job compared to that in the current job; hence, on-the-job search is profitable if and only if \(e_{on} < x_{s}\). Thus, retained workers quit with probability \(I(x_{s})\) while all laid-off workers seek and find employment.\footnote{\(32\)}

As we have emphasized, there are two different ways for workers to leave low-VMP firms, by quits or layoffs. The former has the advantage of selecting workers with the lowest \(e_{on}\) values for search; the latter has the potential advantage of forcing workers to use a relatively more efficient off-the-job search technology and of generating mobility even with close to full insurance. To illustrate these effects and the possibility of an interior layoff solution, we evaluate the derivative of the

\[
\frac{\partial \pi}{\partial (w_{s} - \theta_{s})} = \mu(\theta_{H}/2) - \mu(w_{s} - \theta_{s}/2) > 0.
\]

That is, since \(\mu(0) = 0\), it always pays a laid-off worker to search since \(\mu(\theta_{H}/2) - c < 0\) for all \(c\) in \([0, \mu(\theta_{H}/2)]\). Also, note that, in the present model, lowering wage \(w_{s}\) encourages quits by raising the maximum search effort level \(x_{s}\) consistent with a profitable move; thus, as before, \(q_{s} < 0\) and hence our basic incomplete insurance result, (3), holds here as well.
Lagrangian, \( L = W + \lambda Z \), where
\[
\frac{\partial L}{\partial r} = U(w_L, h_L) - V(0) + \lambda \sigma (w_L, h_L, \theta_L) \\
= - \left[ (1 - I(x_L)) x_L + \int_0^{x_L} e dI(e) \right] \\
+ \varepsilon + \left[ \lambda (1 - I(x_L)) (\theta_L^2 - w_L) \right].
\]

This is a special case of (14a).

If, at one extreme, the off-the-job search is costless (\( \varepsilon = 0 \)), \( \partial L/\partial r \) is strictly negative because the first and last terms are strictly negative (since \( w_L > \theta_L^2 \)) while the middle term is zero; thus all workers are laid off. If, at the other extreme, on-the-job search is costless so that \( I(x_L) = 1 \), \( \partial L/\partial r \) is strictly positive because the middle term is positive while the first and last terms are zero; in this case all workers are retained.

We conclude that preferences and search technologies can always be found such that the equilibrium layoff rate is positive at low-productivity firms. If, in addition, \( d^2 L/d^2 r < 0 \), the equilibrium layoff rate will generally be less than one; that is, some workers are retained and so quits accompany layoffs. (\( L \) will typically be concave in \( r \) because of diminishing marginal utility of income or (while outside the present model) diminishing returns to labor).

**B. Quits Precede Layoffs.** We now illustrate the screening and incentive effects that occur only when quits precede layoffs. The main features of the model in Part A are unchanged except that the \emph{ex post} period is divided into two subperiods of equal length (\( t = 1/2 \)). At the beginning of the first subperiod, if VMP values are realized, all workers are retained, their \( e_{off} \) values are realized, and they search on the job; some of these workers quit to work elsewhere. At the beginning of the second subperiod, a fraction of the remaining workers are laid off and the rest are retained; laid-off workers' \( e_{off} \) values are realized and they search off the job. To simplify we assume that on-the-job search takes place only during the initial subperiod.\(^{33}\)

Under these assumptions, the expected utility of a worker at a low-VMP firm equals
\[
(1 - I(x_L)) \hat{U} + I(x_L) \mu(\theta_H^2/2) - \int_0^{x_L} e dI(e),
\]
where \( I(x_L) \) is the probability that the worker quits,
\[
x_L = \mu(\theta_H^2/2) - \hat{U},
\]
\[
\hat{U} = (1/2) \left[ \mu(w_L - \theta_L^2/2) + \mu(w_L - \theta_L^2) \right] + (1-r)V(0),
\]
\[
V(0) = \mu(\theta_H^2/2) - \varepsilon(x_L).
\]

\( U_L \) is the wage paid to workers who remain with the firm, and \( \delta(x_L) \) is the expected off-the-job search cost of workers whose on-the-job cost exceeded \( x_L \) during the initial subperiod. The corresponding expected profit per worker equals
\[
(1/2)(1 - I(x_L)) \left[ (\theta_L^2 - w_L) + r(\theta_L^2 - w_L) \right].
\]

Differentiation of the corresponding Lagrangian \( L \) establishes that \( w_L = U_L \).

Suppose workers prefer to be retained during the second subperiod, that is, \( \mu(w_L - \theta_L^2/2) > V(0) \) (see Section I, Part C). Ignoring the common factor 1/2 and differentiating \( L \) gives
\[
\frac{\partial L}{\partial r} = (1 - I) \left[ \mu(w_L - \theta_L^2/2) - V(0) + \lambda (\theta_L^2 - w_L) \right] \\
+ \left[ (1 - I)(1 - r) \frac{dV}{dx_L} + \lambda(1 + r) (\theta_L^2 - w_L) \right] \left( - \frac{dI}{dx_L} \right) \frac{\partial x_L}{\partial r}.
\]

This is a special case of (14b). The first term is standard; it equals the expected benefit for workers, \( (1-I)(\mu - V) \), minus the expected cost for firms, \( (1-I)\lambda(w_L - \theta_L^2) \), of a marginal increase in the retention probability \( r \). The second term is more interesting; it measures the net marginal benefit of increasing the retention probability that comes about through the induced change in search.

To sign the second term, observe that increasing the on-the-job search cost below which workers quit, \( x_L \), increases the likelihood that a worker will initially quit; hence \( d^2 x_L/\partial r^2 > 0 \). Observe also that the initial subperiod of employment screens out the lower on-the-job search cost workers; since it is most likely that workers who did not initially quit because of high on-the-job search costs will also experience high off-the-job costs, we assume that \( d\varepsilon /dx_L \geq 0 \) and therefore that \( dV^2/dx_L < 0 \). Finally, since \( \mu > V \), increasing the retention rate increases the payoff, \( r(1-r)V \), to unsuccessful search on the job; hence increasing the retention rate will cause a worker who was initially indifferent to forego search, so that \( \partial x_L/\partial r < 0 \).\(^{34}\) Thus, \( \partial V/\partial r \geq 0 \) and \( \partial I/\partial r < 0 \); in this example, therefore, when on-the-job search precedes layoffs, the screening effect of on-the-job search decreases the layoff rate and the incentive effect of layoffs increases the layoff rate.

\(^{33}\)To make models where layoffs precede and follow search comparable, it should be the case that if retained workers can search \( N \) times in one situation, then they can do likewise in the other situation. In our simple examples, \( N = 1 \).

\(^{34}\)Since \( \mu(\theta_H^2/2) - \hat{U} \) is a positive nondecreasing function of \( x_L \), that intersects \( x_L = \chi_L \) from above, and since raising \( r \) decreases \( \mu(\theta_H^2/2) - \hat{U} \) for each \( x_L \) value, it follows that the solution of \( x_L = \mu(\theta_H^2/2) - \hat{U} \) is a decreasing function of \( r \).
REFERENCES


