NBER WORKING PAPER SERIES

IMPLICIT CONTRACTS, LABOR MOBILITY AND UNEMPLOYMENT

Richard Arnott

Arthur Hosios

Joseph Stiglitz

Working Paper No. 2316

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 1987

The comments of two anonymous referees were extremely helpful. Arnott and Hosios would like to thank the Social Sciences and Humanities Research Council of Canada for financial support, while Arnott and Stiglitz would like to thank the National Science Foundation. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #2316
July 1987

Implicit Contracts, Labor Mobility and Unemployment

ABSTRACT

Firms' inability to monitor their employees' search effort forces a tradeoff between risk-bearing and incentive considerations when designing employment-related insurance. Since the provision of insurance against firm-specific shocks adversely affects workers' incentives to find better jobs, the optimal contract provides only partial insurance: it prescribes low (high) wages and under (over) employment to encourage workers to leave (stay) at low (high) productivity firms; and it employs quits and layoffs as alternative means of inducing separations at low productivity firms, with the mix depending upon the relative efficiency of the on- and off-the-job search technologies.

Our analysis of implicit contracts with asymmetric search information establishes that any consistent explanation for worksharing, layoffs, severance pay, quits and unemployment must focus on questions of labor mobility.

Richard Arnott
Queen's University
Kingston, Ontario
CANADA

Arthur Hosios
University of Toronto
Toronto, Ontario
CANADA

Joseph Stiglitz
Department of Economics
Princeton University
Princeton, NJ 08544

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 in recent years that at least the simpler versions of this theory would not do; though they explained real wage rigidities as a consequence of risk-averse workers demand for insurance against variations in their value of their marginal product, they could not explain unemployment.[1]

To explain unemployment, one must explain, first, why reductions in demand take the form of layoffs rather than reduced hours (worksharing), and second, why those on layoff do not immediately secure employment elsewhere. Traditional implicit contract theory, however, has ignored both issues: it has <u>assumed</u> that reductions in labor demand take the form of layoffs; and it has <u>assumed</u> that all separated workers are immobile, due, say, to prohibitive mobility costs. Indeed, the latter assumption is particularly important as it implies that (i) laid-off workers are necessarily unemployed, (ii) quits are absent, (iii) high-demand firms cannot possibly hire newly separated workers and (iv), when optimally determined severance pay is provided, laid-off workers are actually better-off than retained workers. In fact, of course, worksharing, layoffs, quits and interfirm 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 general theory of implicit contracts which resolves most of these difficulties. Our theory is based on three simple observations: first, the risk which individuals wish to be insured against is not just a decrease in the value of the marginal product (VMP) of the firm at which they are currently employed, and hence a decrease in their wage, but rather, it is a decrease in this value together with the failure to find an alternative job at which their VMP is high; second, finding a job requires search, and search involves costs, though they are generally not prohibitive; and third, wages and severance pay serve a crucial role in this process, by influencing workers' search efforts, in reallocating labor form uses where its VMP is low to those at which it is high.

We follow traditional contract theory in assuming that workers are risk-averse and firms are risk-neutral. Firms thus 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 values of the marginal productivities of labor in different uses) since each worker would move immediately to the job where his VMP is highest. By contrast, if search were costly but perfectly monitorable, firms would provide insurance against relative shocks, but these insurance contracts would specify the intensity of job search.

In fact, not only is search an individually costly activity, but it is also a private activity that is costly to monitor. Indeed, it is frequently impossible to observe either search effort or the outcomes of search. That is, it is difficult for the worker's current employer to monitor outside wage offers or even to verify whether a laid-off worker has been reemployed. In turn, because the employer is unable to observe search inputs and outcomes, the following two related moral hazard problems arise.

The first moral hazard problem results from the unobservability of the best job offer, even when search effort is observable. If both search effort and job offers were observable, the insurance contract provided by the firm would: (a) specify the level of search effort; (b) require an employed (laid-off) worker to move provided he found any job with a wage in excess of his productivity at his current firm (in home production);[3] and (c) compensate the worker with the difference (possibly negative) between his current wage and this alternative wage offer.

When wage offers are not observable, however, this scheme cannot be implemented. The firm can then only provide a fixed severance payment to workers who quit or are laid off, and a given wage to workers who are retained. In designing these payments, however, the following trade-off between risk and efficiency is confronted: the more insurance a firm provides against low-VMP draws, 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. In other words, more efficient quit behavior requires a departure from complete insurance.[4]

The second moral hazard problem results from the unobservability of search effort or intensity. When search effort cannot be monitored and hence cannot be specified contractually, a similar risk-efficiency trade-off presents itself: the more insurance a firm provides against low-VMP draws, the lower the worker's search effort, and hence 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 optimal insurance contract will again provide incomplete insurance.

With private job-search information, the optimal 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.[5] In this setting, moreover, we observe that quits (induced by relatively low wages) and layoffs are simply alternative means of creating separations at low-VMP firms; whereas 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, layoffs have the potential advantage of forcing workers to use a different and likely more efficient off-the-job search technology. As these two instruments are imperfect substitutes, one expects to observe both (i.e. worksharing and quits as well as 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.[6] Regrettably, the simplest model which can be constructed to remedy 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 a partial equilibrium analysis of firms' wages, severance pay and internal distortions when workers' search efforts are private information. Section II then presents a stripped-down general equilibrium model of worksharing, quits, layoffs and unemployment when workers' search costs are private information. 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 is laid off as prescribed by the contract. Subsequently, 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 procede that this is not the only possible interpretation of our model (see footnotes 12 and 13), and that most of our qualitative results, for instance, concerning incomplete insurance, underemployment and the simultaneous use of layoffs and worksharing would remain valid under alternative sequencings as well, e.g. if the firm did not make its layoff decisions until after workers announced whether or not they were quitting (see footnotes 14 and 28).

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.[7]

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, θ_L and θ_H with $\theta_L < \theta_H$. Workers have identical tastes and technical ability. As in 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 $C = \{(w_L, h_L), r, s, (w_H, h_H)\}$. Under this contract employed workers are paid w_i for h_i units of labor when $\theta = \theta_i$; a worker is laid off with probability 1-r in the low-VMP state θ_L; and each laid-off worker is given the fixed non-negative severance payment s.[8]

The terms of employment offered by firms in the ex ante period will depend on what alternative opportunities are available to workers in the 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 $\bar{h} > 0$.[9] Workers are of course concerned with the <u>highest</u> wage they can expect to 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 highest wage offers, given search intensity e, for those workers who are employed and unemployed, respectively; f and g are the corresponding density functions. Additionally, F and G are assumed to satisfy:

$$F(0,0) = G(0,0) = 1;$$

$$F_e, G_e < 0 \quad \text{for} \quad z > 0.$$

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 substitution of the optimal search effort, 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[10]

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

Taking account of workers' optimal search and quit behavior, firms' profits per retained employee are a function of the state θ and of the worker's wages and hours, $\pi(w,h,\theta)$, whereas the cost per laid-off worker is exactly 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

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

The remaining portions of this section basically describe and interpret 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, i.e. when will there be at least partial wage flexibility? When will those who stay with the firm feel underemployed, i.e. when will they wish to work additional hours at the going wage. And finally, if there are layoffs, when will laid-off workers be worse-off than retained workers?

I.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 just the utility he obtains 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 he gets 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:[11]

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

where $x = x(w,h)$ is the wage offer which makes the individual indifferent between quitting and staying, and is defined implicitly by $\alpha(w,h) = \alpha(x,\bar{h})$, $F(x,e)$ is the probability of drawing a wage offer less than x given search intensity e, and the expected utility of a worker who quits is

$$\sigma(x,e) = \left(\int_x^\infty \alpha(z,\bar{h})f(z,e)dz\right)/(1-F(x,e)).$$

We assume that α is a strictly concave function satisfying $\alpha_w>0$ and $\alpha_h<0$ and that β is an increasing convex function of e satisfying $\beta_h\geq 0$ and $\beta(0,h) = 0$. The term $\beta(e,h)$ is meant to capture the cost (in utility terms) of searching at intensity e when the present job demands h units of labor.[12,13]

It follows analogously that the expected utility of a laid-off worker who receives severance pay s and expends e on search is equal to

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

where $y=y(s)$ is the lowest acceptable wage offer and is defined by $\alpha(s,0) = \alpha(y+s,\bar{h})$. 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 respectively described by:

$$e(w,h) = \text{argmax } u(w,h,e),$$

$$e(s) \quad = \text{argmax } v(s,e).$$

It follows that their corresponding equilibrium utility levels are $U(w,h) = u(w,h,e(w,h))$ and $V(s) = v(s,e(s))$, and that firms' equilibrium quit and profit functions per retained worker are, respectively,

$$q(w,h) = 1-F(x(w,h),e(w,h)),$$

$$\pi(w,h,\theta) = (1-q(w,h))(\theta h-w).$$

The quit rate is a function only of the wages and hours offered on the job, and is equal to the probability of finding 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.[14]

I.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 as well.

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

$$\text{(1a)} \qquad \frac{U_w(w_L,h_L)}{U_w(w_H,h_H)} = \frac{\pi_w(w_L,h_L,\theta_L)}{\pi_w(w_H,h_H,\theta_H)} ,$$

$$\text{(1b)} \qquad \frac{U_h(w_i,h_i)}{U_w(w_i,h_i)} = \frac{\pi_h(w_i,h_i,\theta_i)}{\pi_w(w_i,h_i,\theta_i)} \quad i = L,H.$$

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 <u>plus</u> 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' optimal search efforts and quit rates.

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

(2a) $$U_w(w_i,h_i) = (1-q(w_i,h_i))\alpha_w(w_i,h_i),$$

(2b) $$U_h(w_i,h_i) = (1-q(w_i,h_i))\alpha_h(w_i,h_i) - \beta_h(e(w_i,h_i),h_i),$$

(2c) $$\pi_w(w_i,h_i,\theta_i) = -(1-q(w_i,h_i)) - A_i q_w(w_i,h_i),$$

(2d) $$\pi_h(w_i,h_i,\theta_i) = (1-q(w_i,h_i))\theta_i - A_i q_h(w_i,h_i),$$

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

Workers are immobile when they simply choose to forgo search, setting e(w,h)=0 and hence q(w,h)=0. (This outcome is most likely when workers face either a particularly skewed distribution of wage offers or a high marginal disutility from search.) With an immobile workforce 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

$$\alpha_w(w_L,h_L) = \alpha_w(w_H,h_H),$$

$$\frac{-U_h(w_i,h_i)}{U_w(w_i,h_i)} = \frac{-\alpha_h(w_i,h_i)}{\alpha_w(w_i,h_i)} = \theta_i, \qquad i = L,H.$$

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; if either α is additively separable or the supply of labor is inelastic ($h_i = h$), 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. A higher contract wage causes searching workers to become more selective in evaluating job offers. This reduces the expected return to search, and so workers' optimal on-the-job search effort falls. Hence the quit rate falls when the wage is raised, i.e. $q_w(w_i, h_i) < 0$.[15] Thus, making use of the fact that $A_L < 0 < A_H$, (1a) and (2a,c) give

$$(3) \qquad \alpha_w(w_L, h_L) > \alpha_w(w_H, h_H) \ .$$

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 when quits are sensitive to wages.

Since firms are subsidizing their workers in the low-VMP state as part of the insurance package, they would like to encourage as many as possible to search and quit. The problem is that workers who have searched unsuccessfully are indistinguishable from those who make no attempt to search at all. Hence firms must lower the wage to provide appropriate search incentives, even though this entails a departure from full insurance. Conversely, when positive premia are collected in high-VMP states, wages are raised to discourage quits.

In particular, if the utility function α is separable, or labor supply is inelastic, (3) gives $w_H > w_L$. In these circumstances, the model generates a positive correlation between wages and VMP's.[16]

Mobile Labor: Underemployment

We now examine whether our model generates underemployment. Unfortunately, it is not immediately clear in the present search context how this term should be defined. In particular, is there underemployment when

$$-U_h/U_w < \theta \ ,$$

i.e. 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

$$-\alpha_h/\alpha_w < \theta \ ,$$

i.e. when the MRS between wages and hours along the instantaneous utility function is less than the MRT? Reversing these inequalities establishes the corresponding definitions of overemployment that account for and ignore induced search, respectively.

From (2a,b), it is apparent that

$$U_h/U_w = \alpha_h/\alpha_w - \beta_h/(1-q)\alpha_w \ ,$$

and hence these two definitions of underemployment (and of overemployment) coincide if $\beta_h = 0$, i.e. if increasing hours worked does not affect the cost of searching (at any level of search effort). Normally, however, we think of β_h as being positive (increasing hours worked on the job makes search more costly), 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 (using the instantaneous utility function), there is overemployment under the first criterion.

Since it is much easier to establish results concerning the relative magnitudes of $-U_h/U_w$ and θ, we henceforth use the terms under- and overemployment to describe situations where $-U_h/U_w <$ and $> \theta$, respectively; and so production efficiency is said to result when the two are equal. We begin by showing that so long as increasing hours worked reduces quits, there will be underemployment in those states when profits are negative and overemployment in those states when profits are positive. (Analogous results concerning the relative magnitudes of α_h/α_w and θ follow straightforwardly, with some obvious caveats, and are omitted.)

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

$$\frac{-U_h(w_i,h_i)}{U_w(w_i,h_i)} = \frac{(\theta_i - A_i q_h(w_i,h_i)/(1-q(w_i,h_i)))}{(1+A_i q_w(w_i,h_i)/(1-q(w_i,h_i)))} \tag{4}$$

Earlier we noted that increasing wages reduces quit rates. Thus if, in addition, increasing hours also reduces quit rates, so that $q_h < 0$, then

$$\frac{-U_h(w_i,h_i)}{U_w(w_i,h_i)} \gtreqless \theta_i \quad \text{as} \quad A_i \gtreqless 0 . \tag{5}$$

That is, 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.

It is possible, however, that q_h can be positive, even for well-behaved utility functions. On the one hand, an increase in hours worked (at fixed wages) makes the job less attractive, and this increases quit rates, just as an increase in wages (for fixed hours) makes the job more attractive, and this reduces quit rates. On the other hand, however, if search is also time intensive, then a worker who works more has less time to search, and this reduces quit rates.

Returning to (5), observe that its derivation from (4) does not rely upon (2a,b) but depends instead upon specific derivative properties of the quit rate function. However, by taking advantage of the simple utility functions introduced earlier, we can now derive an even stronger result: <u>there will be underemployment (overemployment) in those states where profits are negative (positive) whenever workers' compensated quit derivative with respect to hours is negative</u>. 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

$$\left.\frac{d\pi}{dh}\right|_U = (1-q)\left[\theta - \left.\frac{dw}{dh}\right|_U\right] - (\theta h - w)\left.\frac{dq}{dh}\right|_U = 0 ,$$

where

$$\left.\frac{dq}{dh}\right|_U = \text{compensated quit derivative.}$$

Recalling that

$$\left.\frac{dw}{dh}\right|_U = -U_h/U_w ,$$

and substituting, we can obtain

$$\text{(6a)} \qquad \frac{-U_h(w_i, h_i)}{U_w(w_i, h_i)} = \theta_i - \frac{A_i}{(1-q)} \left.\frac{dq}{dh}\right|_U ,$$

yielding the result that whenever the compensated quit derivative is negative, underemployment occurs in those states where $A_i < 0$ and overemployment occurs in those states where $A_i > 0$. In turn, it can be shown that[17]

$$\text{(6b)} \qquad \left.\frac{dq}{dh}\right|_U = \frac{-B\beta_h}{(1-q)} - \frac{F_e \beta_{eh}}{u_{ee}}$$

where $B > 0$, $F_e < 0$ and $u_{ee} < 0$. Therefore, <u>a sufficient condition in our model for the compensated quit derivative to be negative is that both β_h and β_{eh} be positive, i.e. that increasing hours worked increases both the total and marginal costs of search</u>. (Since

$$q_h = -\alpha_h B - F_e \beta_{eh}/u_{ee} ,$$

where $\alpha_h < 0$, note that (6b) is negative under less restrictive conditions than are required for $q_h < 0$.)

Finally, observe that when workers' labour supply decisions have no effect on their disutility from search, that is, $\beta_h = \beta_{eh} = 0$, then (2a,b) and (6) give

$$\text{(7)} \qquad \frac{-U_h(w_i,h_i)}{U_w(w_i,h_i)} = \frac{-\alpha_h(w_i,h_i)}{\alpha_w(w_i,h_i)} = \theta_i \qquad i = L,H.$$

In this case, a worker's optimal search effort and consequent quit probability depend only upon the level of utility $\alpha(w,h)$, and so do not respond to variations in w and h on a given indifference curve. Therefore, for any given level of utility, and hence given incentive effects, (7) is obviously a necessary condition for profit maximization.

Three general properties of these results are noteworthy. First, unlike most recent implicit contract models with asymmetric information, the occurrence of under- versus overemployment in our model does not rely upon 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 $\alpha = \mu(w-\gamma(h))$.[18,19] Whether there is under- or overemployment depends instead only upon certain weak properties of the utility function ($\beta_{eh}>0$) and upon the contractual availability of insurance for workers ($A_i \neq 0$).

Second, (5) and (6) also indicate that underemployment will occur only in states where workers are being subsidized, with the amount of underemployment being proportional 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. (By contrast, earlier models generated either <u>under</u>employment in all <u>high</u>-VMP states except the highest,[20] or <u>over</u>employment in all <u>low</u>-VMP states except the lowest,[21] and hence predicted either underemployment in (most) good states or overemployment in (most) bad states.)

Finally, (3) and (7) demonstrate that even while employment contracts which are designed to have incentive effects will certainly limit risk-sharing, they need not also generate production inefficiencies.

I.C On Severance Pay

Severance pay is oftentimes portrayed as the Achilles heel of implicit contract theory. The original models largely ignored severance pay, while later work introduced severance pay but obtained the counterfactual result that workers prefer to be laid off rather than retained[22]; that is, contracts which provide full insurance, equating (expected) marginal utilities of income across all states, will result in higher (expected) utilities among those on layoff, in which case we would expect contracts to specify reverse seniority layoff clauses. In this section we show, on the contrary, that under plausible conditions concerning labor mobility and preferences, laid-off workers are indeed worse off than retained workers. Before presenting our results, we review the standard ones.

We begin by taking the layoff rate 1-r as given. The first-order conditions of problem (P) then yield

(8) $$-U_w(w_i,h_i)/\pi_w(w_i,h_i,\theta_i) = V_s(s)$$

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) give

$$\alpha_w(w_L,h_L) > V_s(s) > \alpha_w(w_H,h_H) \quad \text{for} \quad A_L < 0 < A_H,$$

which is the obvious extension of our earlier partial insurance result.

<u>Reverse Seniority</u>

There are two versions of the reverse seniority result. First, suppose employed and laid-off workers are both immobile (don't search) so that (8) becomes

(9) $$\alpha_w(w_L,h_L) = \alpha_s(s,0).$$

The optimal (insurance) contract equates the marginal utilities of income of retained and laid-off workers. Differentiating α_w = constant, gives $dw/dh = -\alpha_{wh}/\alpha_{ww}$, and hence $d\alpha/dh = \alpha_h - \alpha_w\alpha_{wh}/\alpha_{ww}$ for fixed α_w. Therefore, (9) implies that

$$\alpha(w_L,h_L) < (>)\ \alpha(s,0)$$

when leisure is a normal (inferior) good; and hence <u>immobile workers prefer to be laid off when leisure is normal</u>.

Second, suppose employed workers are again immobile, and that laid-off workers are now mobile but face uncertain job opportunities. To simplify, suppose we have a separable utility function

$\alpha(w,h)=\mu(w)-\gamma(h)$. In this case, (8) again yields the result that retained and laid-off workers' (expected) marginal utilities of income are equal, i.e.

$$\mu'(w_L) = \mu'(s)G(y,e) + \int_y \mu'(s+z)g(z,e)dz \quad (10)$$

where $e = e(s)$. It can be shown that (10) implies that $\mu(w_L)$ will be greater than (equal to, less than)

$$\mu(s)G(y,e) + \int_y \mu(s+z)g(z,e)dz$$

as $\mu(\ )$ is an increasing (constant, decreasing) absolute risk-aversion utility function.[23] Thus, <u>whenever search entails zero disutility and hours worked is not a variable, workers with decreasing absolute risk-aversion utility functions will prefer layoffs with severance pay</u>.

<u>Costly Search</u>

Once we recognize, however, that search is costly, the presumption that laid-off workers are better-off no longer obtains. There are two reasons for this. First, if search is costly, and these search costs do not enter the utility function additively with income, then providing income insurance (by equating expected marginal utilities) may not fully compensate for the costs and risks of search; in this event we say that search is (partially) non-pecuniary. Secondly, retained workers will, we have argued, have their hours reduced to encourage them to search more; in particular, hours may be reduced below the "standard" hours at alternative employment, in which case even with full income insurance, their expected utility will be higher than those laid off.

To see these results, first suppose every laid-off worker can effortlessly (with $e=0$) secure a new job which pays z for $\bar{h}$; since there are no incentive problems here, the optimal severance payment will satisfy

$$(11a)\qquad \alpha_w(w_L,h_L) = \alpha_s(s+z,\bar{h}).$$

It then follows that laid-off workers will be worse-off even though leisure is normal whenever the spot job entails more work ($\bar{h} > h_L$). Now, suppose this alternative job can be secured only by exerting some fixed effort $e > 0$; assuming that e is not prohibitive, retained workers are better-off when

$$(11b)\qquad \alpha(w_L,h_L) > \alpha(s+z,\bar{h}) - \beta(e,0).$$

Thus, whenever search costs are in part nonpecuniary, it is clear that (11) can be satisfied without having to impose either the inferiority of leisure or $\bar{h} > h_L$.

Thus, provided search costs are non-pecuniary (and accordingly, severance payments cannot be used to compensate them), and provided there are significant probabilities of being re-employed elsewhere at jobs requiring at least as much labor as at their ex ante firms, then laid-off workers will be worse-off than retained workers; and this will be true even in the presence of decreasing absolute risk aversion. Moreover, when preferences are nonseparable and workers search on the job as well, it is clear that (8) says very little about the relative magnitudes of $U(w_L,h_L)$ and $V(s)$.

I.D <u>Severance Pay for Retained Workers</u>

Our results are not substantially altered if firms can 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 search, firms will still wish to reduce hours of retained workers in the face of a reduction in the demand for their products; so long as reducing wages has some beneficial effects in inducing search, they will not provide complete income insurance to the retained workers; so long as search costs are not completely additive with income, the severence pay will not fully compensate for effort expended on search; and finally, so long as there is imperfect information about the job offer which the individual accepts, the distortions we described earlier with respect to quit behavior, (5) and (6a), will remain.[24]

II. Layoffs and Worksharing[25]

We now determine whether firms' layoff rates are ever positive. Since workers in our model are technically identical, and as hours per worker and the number of workers are perfect substitutes in production, differential productivity and technological arguments for layoffs have been ruled out.[26] While these considerations are not unimportant, we set them aside here to specifically study the impact of costly mobility on layoffs. In particular, we will show that the availability of alternative employment for laid-off workers introduces a natural non-convexity into firms' optimization problems and, as a result, worksharing will generally not dominate layoffs.

We are going to proceed in two stages. First, we will construct a model in which retained workers do not quit, and in which layoffs are the only ex post source of mobility. Then, in Sections II.C and II.D, we will construct models with random search costs in which there are both quits and layoffs.

To simplify the remaining discussion, we henceforth assume that workers' preferences satisfy $\beta(e,h)=e$; with $\beta_h=0$ we know, from (7), that long-term and spot employment will both involve efficient production satisfying $-\alpha_h/\alpha_w=\theta$.

II.A Layoffs Only

We begin by considering a situation in which there is no moral hazard problem because search effort and outcomes are both observable, and in which retained workers do not search. In this case, the optimal employment contract fully insures retained workers (who don't quit), and hence satisfies $\alpha_w(w_L,h_L)=\alpha_w(w_H,h_H)$ and $-\alpha_h(w_i,h_i)/\alpha_w(w_i,h_i) = \theta_i$; thus, with well-behaved utility functions, employees will work more in high-productivity states.

To easily demonstrate that worksharing and layoffs are not mutually exclusive outcomes, it will be useful at this juncture to specialize the model further by assuming that the utility function takes on the special form $\alpha(w,h)=\mu(w-h^2/2)$. We shall refer to $w-h^2/2$ as the "net income" of a job, that is, the wage income net of labor costs. Given $-\alpha_h/\alpha_w=\theta$, we can immediately determine that workers will supply $h_i=\theta_i$ hours at firms where $\theta=\theta_i$. Let $y_i = \theta_i^2$, $i = L,H$, denote the output, $\theta_i h_i$, of a worker in state i , so that in equilibrium $\alpha(w_i,h_i)=\mu(w_i-y_i/2)$.

For simplicity, we also consider the special case of the search process in which a worker either expends e to become fully informed of all employment opportunities, or forgoes search and remains uninformed (and immobile). We assume that the requisite effort level e is fixed and that it always pays an unemployed worker to search (i.e. $0 < e < \mu(y_H/2)$).

Now, with firm-specific shocks and constant returns, there will always be a group of high-VMP firms who are 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 $z=y_H$ for $\bar{h} = \theta_H$ units of labor. Observe that $\alpha(z,\bar{h}) = \mu(y_H/2)$.

We can show that, in equilibrium, retained workers do not search (quit) while every laid-off worker searches. Therefore, retained and laid-off workers' expected utilities are described, respectively, by:

$$U(w_i,h_i) = \mu(w_i - y_i/2) \quad i=L,H,$$
$$V(s) = \mu(s+y_H/2)-e.$$

The equilibrium contract will therefore maximize expected utility

$$W = r\mu(w_L-y_L/2) + (1-r)(\mu(s+y_H/2)-e) + \mu(w_H-y_H/2)$$

subject to the zero expected profit constraint

$$Z = r(y_L-w_L) - (1-r)s + (y_H-w_H) = 0.$$

It should be clear that employed workers will receive full insurance, that is, $\mu'(w_L-y_L/2) = \mu'(w_H-y_H/2)$. If, as we assume here, negative severance payments are not permitted, it can be shown that the equilibrium value of s must be zero.[27] (Negative values are examined later below.)

The model's remaining equilibrium properties are derived as follows: Solving $Z=s=0$ and $w_L-y_L/2 = w_H-y_H/2$ gives

$$w_i - y_i/2 = (ry_L + y_H)/2(1+r) = \delta(r) ,$$

so that workers' expected utility W can be rewritten as

$$W(r,e) = (1+r)\mu(\delta(r)) + (1-r)(\mu(y_H/2)-e) .$$

Observe that $W_{rr} < 0$ for $0 \le r \le 1$. There are two corner solutions to consider: if $W_r = 0$ at $r = 0$, no one is retained; and if $W_r = 0$ at $r = 1$, everyone is retained. To determine when these solutions occur, we define the critical effort levels, $e_u(y_L)$ and $e_l(y_L)$ by the first-order conditions

$$W_r(1, e_u(y_L)) = 0 = W_r(0, e_l(y_L)) .$$

It can be shown that these effort levels are both decreasing functions of y_L (and hence θ_L) satisfying $\mu(y_H/2) > e_u(y_L) > e_l(y_L) > 0$ for $y_L < y_H$ and $e_u(y_H) = e_l(y_H) = 0$ for $y_L = y_H$. Thus, if $e \ge e_u(y_L)$, $r = 1$; if $e \le e_l(y_L)$, $r = 0$; and if $e_l(y_L) < e < e_u(y_L)$, $0<r<1$ where $dr/de > 0$ and $dr/d\theta_L > 0$. See Figure 2.

The trade-off between risk-sharing considerations and search costs is clear. When either search becomes less expensive or the low-VMP insurance subsidy becomes larger (as θ_L falls), the balance tips towards layoffs and search, and away from worksharing. For example, any combination of e and y_L below (above) the shaded area in Figure 2 involves layoffs (worksharing) and no worksharing (layoffs). These corner solutions are familiar: when workers are freely mobile (e=0), one-period implicit contract models must replicate spot auctions; and when workers are immobile ($e>\mu(y_H/2)$), the concavity of μ ensures that worksharing strictly dominates layoffs. <u>For intermediate values however, only some workers are retained to workshare in the low-VMP state while the remainder are laid-off with zero severance pay and undertake search</u>.

II.B On Severance Pay (Again)

With constant returns in production and identical workers, one might initially conjecture that the equilibrium employment contract will treat all workers identically ex post, and hence that the optimal layoff rate must be a corner solution, either zero or one. In fact, of course, interior solutions satisfying $0 < r < 1$ are indeed possible, as demonstrated above. We begin this section by asking whether this result is due to our earlier prohibition of negative severance pay.

Interior Solutions

The analysis below considers the more general case in which laid-off workers receive job offers with net income $z/2$, where z need not equal y_H. The equilibrium contract with unrestricted severence pay is then derived as follows. Solve the zero-profit constraint, $Z = 0$, and the first-order conditions for full insurance, $w_i - y_i/2 = s + z/2$ $(i = L, H)$, to obtain

$$w_i - y_i/2 = s + z/2 = (ry_L + y_H + (1-r)z)/4 = \rho(r) \; ;$$

then, substituting these expressions into W, we get the expected utility function

$$W(r,e) = 2\mu(\rho(r)) - (1-r)e \; .$$

Since $W(r,e)$ is a strictly concave function of r, we can proceed as before to verify that an interior solution will be optimal for some values of (y_L, z, e). For example, if $z > y_L$,

(12) $$W_r = \mu'(\rho)(y_L - z)/2 + e \; ,$$

which shows that there always exists an open interval of effort levels consistent with interior optima. Therefore, while negative severance pay may pose enforcement problems, it certainly does not preclude interior layoff solutions.

The cost of laying off a worker to a firm is the severance payment s plus the forgone profit $y_L - w_L$. Since $w_L - y_L/2 = s + z/2$, this cost equals

$$s + (y_L - w_L) = (y_L - y_L/2) - z/2 .$$

Therefore, layoffs are profitable whenever the worker's net outside income, $z/2$, exceeds his competitive internal net income, i.e. his VMP minus the disutility of work, $y_L - y_L/2$. In turn, when layoffs are profitable, increasing the layoff rate increases employed workers' net income $\rho(r)$. Therefore, interior solutions are possible in this model with complete insurance because, when $z > y_L$, the marginal benefit of layoffs $\mu'(\rho)(z-y_L)/2$ is a positive but continuously decreasing function of the layoff rate (due to risk aversion), while the marginal search cost e takes on positive fixed values. More generally, <u>interior solutions are possible with complete insurance whenever search involves some non-pecuniary and hence uninsurable costs</u>.

Notice, however, when search is costless, $e = 0$ and (12) predict the corner solutions $r = 1$ when $y_L - y_L/2 > z/2$ and $r = 0$ when $y_L - y_L/2 < z/2$; in addition, when laid-off workers' utility is alternatively represented by $\mu(s+z/2-e)$, so that all search costs are pecuniary, interior equilibria are again ruled out as $r = 1$ when $y_L - y_L/2 > z/2 - e$ and $r = 0$ otherwise. Therefore, with complete insurance and only pecuniary search costs, the equilibrium layoff rate simply implements the standard "first best" allocation.

Production Efficiency

It has been widely thought that the absence of severance pay is the only reason that layoffs are not at a productively efficient level. Without severance pay, firms can only provide insurance by retaining and subsidizing workers even when their VMP is higher elsewhere. We now show that even with optimally chosen severance pay, the layoff rate will not generally be productively efficient. There are two parts to the argument: first, an efficient layoff policy requires complete insurance, not just the availability of (negative) severance pay; and second, the moral hazard problems described in Section I preclude complete insurance so that (as with quit behavior) production inefficiencies will again result.

To see this, consider a set-up similar to Section I in which, given e, a laid-off worker's best alternative option, yielding $z/2$, is a random draw from a nondegenerate distribution with c.d.f. $G(z/2,e)$, where $G(.,0) = 1$. Letting $s(z)$ denote the severance payment to workers who draw $z/2$, the first-order condition for layoffs, at an interior solution, yields (after some manipulation)

$$y_L = w_L - E(s(z)) - Q \tag{13}$$

$$Q = [\mu(w_L - y_L/2) - (E\mu(s(z)+z/2) - e)]/\mu'(w_L - y_L 2) \ .$$

Hence the output gain from retaining an extra worker, his VMP, equals the corresponding marginal cost of an extra employee, that is, the contract wage minus the expected severance pay minus the expected (implicit) insurance indemnity. There are three cases of interest.

First, suppose search effort is fixed and firms can monitor the outcomes of search, i.e. each laid-off worker's $z/2$ realization is observable. In this case, the optimal severance payment $s(z)$ satisfies $\mu'(w_L-y_L/2) = \mu'(s(z)+z/2)$ for all z. Therefore, $w_L-y_L/2$ equals $E(s(z))+E(z/2)$, $\mu(w_L-y_L/2)$ equals $E\mu(s(z)+z/2)$, so that $Q= e/\mu'(\rho)$, and hence

$$y_L-y_L/2 = E(z/2) - e/\mu'(\rho) .$$

In words, the net income from current employment equals the expected net income from alternative employment minus the cost of search, and therefore, the layoff rate is efficient when firms provide complete insurance.

Second, suppose search effort is fixed but firms cannot monitor workers' job offers. In this case, $s(z)$ is the payment to workers who claim to draw $z/2$, and the only incentive compatible payment schedule is the fixed non-contingent payment, $s(z) = s = E(s(z))$. Since $E\mu(s+z/2)$ is strictly less than $\mu(s+E(z/2))$,

$$Q > (w_L-y_L/2) - (s+E(z/2)) + e/\mu'(\rho) ,$$

and so,

$$y_L-y_L/2 < E(z/2) - e/\mu'(\rho) . \tag{14}$$

Therefore, when success in job search cannot be monitored, firms overemploy workers in the low-VMP state, that is, retained workers' net current income is strictly less than their expected net alternative opportunity.

While this distortion obviously also results when severance pay is prohibited $(s=0)$, earlier research had mistakenly focused on the absence of severance pay per se, rather than the absence of complete insurance, as the source of the problem. We therefore conclude that the allocative impact of severance pay vis-a-vis layoff rates is minimal when, as seems plausible, workers' outside opportunities are uncertain and unobservable: layoff rates are certainly higher when severance pay is provided, but are still deficient relative to the productively efficient solution.

Finally, suppose search effort is variable but unobservable, while search outcomes can be observed. In this case, the firm can provide complete insurance in the spot market by offering severance pay which satisfies $\mu'(s(z)+z/2) = \text{constant}$, for all z. However, because search entails disutility, when $s(z)+z/2$ is constant, the optimal search effort will be zero and hence workers remain immobile. To overcome this moral hazard problem and encourage search, an alternative schedule of payments will be offered which, at a minimum, involves nonconstant $s(z)+z/2$. Therefore, complete insurance though feasible is not desirable, and overemployment, as in (14), results in the low-VMP state.

II.C Layoffs With On-the-Job Search

The model of section II.A was structured so that there was no on-the-job search (and hence no quits). Thus, the model could not accurately capture the trade-off between inducing voluntary separations (through lower pay to retained employees) and involuntary separations (through layoffs). To proceed in the simplest way, we extend the model in Section II.A by assuming search costs are random; there is some probability that search costs are low (here, for simplicity, assumed to be zero), and some probability that search costs are high (here, for simplicity, assumed to be sufficiently high that individuals would in fact never search); and neither the worker nor the firm know what their search costs will be before the contract is signed (and indeed, before the layoff decision is made). In particular suppose

$$\text{on-the-job} \begin{cases} \text{Prob}(e=0) = p, \\ \text{Prob}(e=\mu(y_H/2)) = 1-p \ ; \end{cases}$$

$$\text{off-the-job} \begin{cases} \text{Prob}(e=0) = q, \\ \text{Prob}(e=\mu(y_H/2)) = 1-q. \end{cases}$$

We expect that $p < q$ as laid-off workers generally access a better job-finding technology than retained workers. Given these extreme distributions, it can be shown that firms will have no incentive to offer contracts designed to discourage quits by highly mobile workers at high productivity firms or encourage quits by immobile workers at low productivity firms.

From our earlier analysis, we expect the optimal severance payment to be a decreasing function of q, which measures the degree of mobility available to laid-off workers. Here we simplify by setting $s=0$. The resulting equilibrium layoff-worksharing mix corresponding to any θ_L and combination of job-finding technologies is depicted in Figure 3; all points to the left (right) of the 3-D wedge yield contracts with layoffs (worksharing) and no worksharing (layoffs). When off-the-job search is relatively efficient (high q/p), layoffs become the primary vehicle for inducing separations; and when on-the-job search is relatively efficient, worksharing with quits becomes more attractive. In fact, layoffs can occur only if off-the-job search is more efficient, i.e. $q>p$.

II.D Random Search Costs

In Section II.C, the optimal layoff rate was determined by the relative efficiency of workers' on- and off-the-job search technologies, whereas wage-hours policy had no effect on quits. By contrast, our earlier analysis in Section I ignored layoff rates and argued that by lowering wages (and hence reducing the effective degree of insurance) firms could induce quits and hence movement from low-productivity to high-productivity firms. This section shows how the firm must balance off simultaneously the effects of all of its terms of employment, that is, of wages and hours, layoff rates, and severance pay.

The model is the same as that of the previous section, except that now there is a continuous distribution of search costs. The consequence of this is that as the terms offered retained workers become worse, more of the retained workers are induced to search. In particular, we now suppose retained workers' requisite e's are drawn from a distribution with continuous c.d.f. F(e), that unattached workers' values are drawn from a distribution with continuous c.d.f. G(e), and that F and G share the common support $[0,\mu(y_H/2)]$.

Assuming that retained workers search to quit (rather than quit to search) and substituting the efficient labor supplies $h_i = \theta_i$, agents' expected utilities are again described by W(C) and Z(C) as in Section I, except that

$$U(w_i,h_i) = (1-F(\chi_i))\mu(w_i-y_i/2) + F(\chi_i)\mu(y_H/2) - \int_0^{\chi_i} e\,dF(e),$$

$$V(s) = (1-G(\zeta))\mu(s) + G(\zeta)\mu(s+y_H/2) - \int_0^{\zeta} e\,dG(e),$$

where $\chi_i = \mu(y_H/2) - \mu(w_i-y_i/2)$ and $\zeta = \mu(s+y_H/2) - \mu(s)$, and

$$\pi(w_i,h_i,\theta_i) = (1-F(\chi_i))(y_i-w_i).$$

χ_i is the difference in utility between what the worker gets at his job and what he would get if he is successful in finding a new job; hence search is profitable if and only if $e < \chi_i$. Similarly, ζ is the difference between the utility of a laid-off worker who searches and finds a job (exclusive of search costs) and his utility if he does not search; hence laid-off workers search if and only if their search costs are less than ζ. Thus, retained workers quit with probability $F(\chi_i)$ and laid-off workers seek employment with probability $G(\zeta)$.

In our earlier model, lowering wage w_L encourages quits by increasing search effort and hence increasing the likelihood that a random outside offer will dominate w_L; in the present model, lowering w_L raises the maximum search effort level χ_L consistent with a profitable move, and hence increases the likelihood that a randomly drawn search cost will be less than χ_L.

There are, as we have emphasized, 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 values for search, while the latter has the potential advantage of forcing workers to use a relatively more efficient off-the-job search technology. To illustrate these effects and the possibility of an interior layoff solution, we evaluate the derivative of the Lagrangian, $L = W+\lambda Z$, where

$$\frac{\partial L}{\partial r} = U(w_L,h_L) - V(s) + \lambda(\pi(w_L,h_L,\theta_L) + s) \tag{15}$$

$$= - [(1-F(\chi_L))\chi_L + \int_0^{\chi_L} edF(e)] + [(1-G(\varsigma))\varsigma + \int_0^{\varsigma} edG(e)]$$

$$+ [\mu(y_H/2) - \mu(y_H/2+s) + \lambda s] + [\lambda(1-F(\chi_L))(y_L-w_L)].$$

To compare search technologies, we now identify one technology as more efficient than another whenever the former has a lower mean effort level.

Thus, taking $F(\ )$ as given, there exists a sequence of increasingly efficient off-the-job search technologies such that, in the limit, $(s,G(\varsigma))$ approach $(0,1)$, the 2nd and 3rd bracketed expressions in (15) go to zero and $\partial L/\partial r$ becomes strictly negative (as $s = 0$ implies $y_L-w_L\leq 0$). On the other hand, taking $G(\ )$ as given, there exists a sequence of increasingly efficient on-the-job search technologies such that, in the limit, $F(\chi)$ approaches one for all $\chi > 0$, the 1st and 4th bracketed expressions in (15) go to zero and $\partial L/\partial r$ becomes strictly positive (as the f.o.c. $\lambda \geq V_s(s)$ implies $\lambda \geq \mu'(y_H/2)$).

It follows that preferences and search technologies can be found such that the equilibrium quit rate, layoff rate and severance payment are all non-zero at low-VMP firms. In these circumstances, the equilibrium rate of unemployment is $(1-r)(1-G(\varsigma))/2$ where $(1-r)/2$ is the workforce proportion on layoff and $1-G(\varsigma)$ is the proportion of unattached workers who draw search costs exceeding $\varsigma = \mu(s+y_H/2)-\mu(s)$, and who therefore have no incentive to search.[28]

In the absence of ex ante employment contracts, the equilibrium unemployment rate here would be zero as $G(\mu(y_H/2)) = 1$. Therefore, implicit contracts and asymmetric job-market information are necessary conditions for a positive equilibrium unemployment rate. In fact, implicit contracts can only generate unemployment by explaining either why laid-off workers are unwilling to search for jobs at high productivity firms (or are unsuccessful), or why the latter firms are unwilling to hire workers laid off elsewhere. We have pursued the former approach.

II.E A Further Comparison Between Layoffs and Quits

To induce search, one needs to impose some risk on some individuals. This risk can take one of two forms: it can be a risk imposed on a few individuals in the form of layoffs, or it can be a risk imposed on all individuals in the form of lower wages to induce on-the-job search. Neither form in general dominates the other.

If firms are imperfectly informed concerning the search abilities, costs, and quit propensities of different individuals, layoffs will not discriminate between efficient and inefficient searchers. With on-the-job search, those who have a comparative advantage in search will do so. The firm may be able to induce all the separations it desires by lowering the wage only slightly to induce those with very low search costs to find employment elsewhere.

(The assumption that the firm is completely uninformed about comparative search costs is not, however, completely correct. Younger workers, for instance, and those who have recently been searching (the newly hired who already have accumulated considerable information about the labor market) may have a relatively more efficient search technology;

hence these are the workers which the firm should layoff. This is, of course, consistent with observed patterns of layoffs.)

On the other hand, those who leave the firm voluntarily may be those who are the firm's most productive workers; that is, the firm may not be able to discriminate between high and low productivity workers as well as other firms may, or alternatively, those with lower search costs may also be significantly more productive. In either event, when the firm lays off workers, it loses a cross-section of its labor force, but when the firm lowers its wage, it loses its best workers. These standard efficiency-wage arguments also rely upon incomplete information and explain why firms may be reluctant to lower the wages of retained workers (see Stiglitz, and Weiss).

III. Concluding Remarks

A complete theory of unemployment should answer the following questions: How do we explain the degree of observed wage flexibility?[29] How do we explain the form of unemployment, e.g. worksharing 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?

Existing implicit contract models, while they provide explanations for some of these phenomena, fail to provide explanations for 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; e.g. it can explain partial wage insurance, internal distortions in the form of underemployment at low-VMP firms, layoffs and quits, sluggish interfirm mobility and equilibrium unemployment.[30]

In this paper we have presented a positive analysis of implicit contracts with asymmetric search information. A normative analysis would be straightforward. From the literature on equilibrium with incomplete markets and imperfect information, it should be apparent 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 in fact generally not constrained Pareto efficient.[31]

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 worksharing, 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 simply not an artifact of the artificial restrictions one has imposed, e.g. that there is no severance pay, no on-the-job search, etc.

Our model can be viewed as an extension of the standard theory to take into account the fact that what is at risk is not being laid off, 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.

Footnotes

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

2. Though more recent developments in implicit contract theory, particularly those based on asymmetries of information (Azariadis, 1983; Sandford Grossman and Hart), 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. In addition, these new theories have also been criticized for their information assumptions, both that the contracts do not employ all the relevant and easily available information, and that they sometimes require the use of information (eg. restrictions on the sale of assets) which is not easily monitored. For a fuller discussion, see Stiglitz (1986).

3. A first best (full information) socially optimal insurance contract would provide that he move whenever he finds a job with a productivity in excess of his current productivity.

4 The same moral hazard problem is examined by Arnott and Stiglitz (1985) in their analysis of implicit contracts when unobservable non-pecuniary attributes are associated with different firms.

5. Note that asymmetric information plays a key role in our analysis, but that the particular source of informational asymmetry is different from that discussed, say, by Azariadis (1983) or Grossman and Hart (1981). In these papers workers cannot observe realizations of firms' revenue functions, which leads to an adverse selection problem. In the present paper, as in Jon Strand's, firms cannot observe workers' search efforts, which leads to a moral hazard problem. The Azariadis-Grossman-Hart kind of asymmetry can be introduced here without substantially altering our results.

6. Related papers which also highlight labor mobility include those by John Geanakoplos and Takatoshi Ito (1981) and Hosios (1985,1986); these papers develop simple general equilibrium mdoels of layoffs, hiring and unemployment in which the ex post job-finding probability is endogenous and, as here, laid-off workers' job offers are private information.

7. 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 an informational advantage over external agents, so that job-related insurance is provided to workers by employers rather than insurance companies.

8. Observe that asymmetric information has ruled out compensation conditional upon a worker's search effort or outside job offer. Severance pay for workers who quit is prohibited here only to simplify the discussion. See Section I.D.

9. This assumption could, of course, easily be removed at some increase in notational complexity. We again emphasize that we are concerned with constructing the simplest model illustrating the points at issue, rather than the most complete.

10. Workers are assumed to not quit to search, i.e., $U(w_i, h_i) \geq V(0)$. This constraint is assumed to be satisfied, and will be ignored in the subsequent discussion.

11. The additively separable structure and the imposition of $\beta_w = 0$ simplify the analysis but are otherwise inessential.

12. More than one interpretation of the model is possible, and each involves a slightly different rationale for the dependency of the search disutility $\beta(e,h)$ on h. For example, it is possible to think of each period being divided into two parts: a job-search period and a production period. Workers' current employers in this view are required to tell workers at the beginning of the job-search period whether or not they are assured a job with that employer during the production period; and so contract design will have to account for the effect of this announcement on workers' behavior during the initial search period.

    In the subsequent analysis, we shall argue that an increase in hours worked increases the cost of search for retained workers. The worker presumably has less time for search (or other non-work) activities. In the interpretation of the model just offered, however, with search preceeding production, it is the substitutability of leisure during the search and production sub-periods that results in the quit rate declining as hours worked increases. An alternative interpretation, in which search and production occur together during an initial sub-period, and where hours worked directly affects search disutility, is formalized in footnote 13.

    Under either interpretation, however, what is crucial is that workers believe that firms will not re-employ laid-off workers who are unsuccessful in finding alternative employment. This commitment is essential if the layoff announcement is to have the desired on-the-job search incentive effects and if the actual layoff is to have the desired off-the-job search incentive effects. Even if workers believe there is a probability, greater than zero but less than one, of the firm reneging on its commitment to not rehire laid-off workers, there will still remain an attenuated search incentive effect.

13. We now formulate a slightly different interpretation of the events we have in mind to show that whether search precedes or is concurrent with production is not critical here.

Suppose the unit period is divided into 2 subperiods, of lengths λ and $1-\lambda$, such that workers supply labour and search during the 1st subperiod and either quit of stay at the beginning of the 2nd subperiod, and therefore

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

Then, defining $\hat{F} = \lambda+(1-\lambda)F$ and $\hat{\beta} = \lambda\beta$, we get

$$u(w,h,e) = \alpha\hat{F} + \sigma(1-\hat{F}) - \hat{\beta}, \text{ where } \hat{F}(0,0) = 1 .$$

That is, except for the "^", this is identical to the expression for $u(w,h,e)$ in the text, and hence our analysis of search and quits goes through in exactly the same way.

14. After firms' layoff decisions have been made, retained workers quit the pool of employed workers with probability $1-F$, and laid-off workers quit the pool of unemployed workers with probability $1-G$. This symmetrical structure simplifies the analysis. Nevertheless, it should be noted that whether layoffs precede or follow quits on-the-job would play a somewhat more important role when there are small firms with

diminishing returns technologies; then, because of the stochastic nature of quits, the firm would be uncertain about the number of employees during the production period, and, because of diminishing returns, variability in employment (in a given state) is costly.

The order would also be of some importance if there were adverse selection effects. As we have emphasized, quits have the advantage of encouraging those with lower search costs to engage in more turnover. There are thus some efficiency gains to having quits precede layoffs. On the other hand, when quits precede layoffs, increasing the layoff rate has positive on-the-job search incentive effects; moreover, our argument for layoffs in Section II, on the grounds that off-the-job search is generally more efficient than on-the-job search, goes through independent of the order of quits and layoffs.

Our central concern in this paper is to establish, under fairly general conditions, that optimal contracts will entail a mixture of quits and layoffs. This central result will remain valid regardless of the sequencing of quits and layoffs; making quits precede layoffs would require a 3-period analysis and would further complicate what is already an admittedly complex setup.

15. $q = 1-F(x,e)$ gives $q_w = -(F_x x_w + F_e e_w)$ where $F_x > 0$ and $F_e > 0$. From the definition of $x(w,h)$, we have $x_w = \alpha_w(w,h)/\alpha_w(x,\bar{h}) > 0$; and from the f.o.c. $u_e(w,h,e) = 0$, where

$$u_e = \alpha(w,h)F_e(x,e) + \int_x \alpha(z,\bar{h})f_e(z,e)dz-\beta_e(e,h)$$

and $x = x(w,h)$, we have $e_w = -\alpha_w(w,h)F_e/u_{ee}$. Assuming that the second-order condition $u_{ee} < 0$ is satisfied, we then have $q_w = -\alpha_w B$ where $B = (F_x/\alpha_x) - F_e^2/u_{ee} > 0$.

16. Previous work has recognized that wages will likely be raised to discourage quits when profits are high or turnover is costly (e.g. Steve Salop (1979)), but fails to make the symmetrical argument when profits are low. Separability is obviously sufficient but not necessary for this correlation between real wages and VMP's.

17. Proceeding as in footnote 15, it can be shown that $q_h = -\alpha_h B - (F_e \beta_{eh}/u_{ee})$, where $B = -q_w/\alpha_w$. Also, from (2a,b),

$$\frac{-U_h}{U_w} = \frac{-\alpha_h}{\alpha_w} + \frac{\beta_h}{\alpha_w(1-q)} = \left.\frac{dw}{dh}\right|_U ,$$

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

$$\left.\frac{dq}{dh}\right|_U = q_h + q_w \left.\frac{dw}{dh}\right|_U$$

gives (6b).

18. While the discussion is often couched in terms of moral hazard, c.f. Grossman and Hart, the question with which the earlier models with asymmetric information are concerned is the inability of at least one party to the implicit contract to distinguish among states of nature (an adverse selection issue); the question with which we are concerned is the inability of firms to monitor workers' search intensities (a moral hazard issue).

19. See Russell Cooper (1983), Hart, and Stiglitz for a description of the role of preferences in generating production inefficiencies.

20. See Azariadis (1983), and Grossman and Hart.

21. See V.V. Chari (1983), Cooper, and Jerry Green and Charles Kahn (1983).

22. See Rosen, and Stiglitz.

23. Whether the expected utility of income is greater or less for the laid-off (vs. retained) worker 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 Harua Imai, Geanakoplos and Ito (1981).

24. 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 $\alpha(w,h) = \alpha(x,\bar{h})$. As a result, $u(w,h,e)$ is replaced by

$$u(w,h,e,b) = \alpha(w,h)F(x-b,e) + \sigma(x,e,b)(1-F(x-b,e)) - \beta(e,h) ,$$

$$\text{where } \sigma(x,e,b) = \int_{x-b} \alpha(z+b,\bar{h})f(z,e)dz/(1-F(x-b,e)) .$$

Hence the optimal search intensity and resulting quit rate are given by $e(w,h,b) = \text{argmax } u(w,h,e,b)$, $q(w,h,b) = 1-F(x-b,e(w,h,b))$, where $q_b > 0$. Similarly, the resulting utility and profit functions are given by

$$U(w,h,b) = u(w,h,e(w,h,b),b),$$

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

As U_w and U_h are again described by (2a,b), and as π_w and π_h are again described by (2c,d) where $A = \theta h-w+b$, it follows that 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.

25. 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 subsection II.A. resembles Baily's (1977).

26. See Ken Chan and Yannis Ionannides (1982), Mark Lowenstein (1983), Rosen, and Andrew Weiss (1980).

27. Since $\mu(y_H/2)-e > \mu(0) = 0$, it turns out that every laid-off worker will search when $s = 0$. To verify that $s = 0$, note that $w_L-y_L/2=w_H-y_H/2$ and $y_H > y_L$ imply $y_H-w_H > y_L-w_L$, so that $Z=0$ and $s\geq 0$ give $y_H-w_H \geq 0$. 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 $\theta_h h_H-w_H=t$, for some $t \geq 0$, $s > 0$ implies $\mu(s+y_H/2) > \mu(w_H-y_H/2)$. It follows that a non-positive s value is required to equate retained and laid-off workers' marginal utilities of income.

28. It should be noted that allowing severance pay for retained workers who quit will effectively make the on-the-job search technology relatively more efficient than otherwise. Hence the optimal layoff rate will be lower, but generally still positive. Also, allowing quits to precede layoffs, to screen out low search cost workers, would have a similar effect on layoffs.

29. The earlier implicit contract models probably explained too much: they suggested that there would be guaranteed annual incomes (in the case of separable utility functions). In fact, as we do observe some wage and income flexibility, the question is, how do we explain the extent of this flexibility, not its complete absence.

30. The assumption of limited information concerning employment status following a layoff has, in fact, been implicit in the previous implicit contract literature. In those models, if workers' employment status could be monitored, all workers would prefer to be laid off, provided the probability of not getting a job was high enough (and provided leisure is a normal good). The "lucky" workers would be those who fail to obtain jobs, and who are thereby fully compensated, and so the smaller the likelihood of obtaining a job, the more would workers clamor to be those laid off.

31. The term constrained Pareto efficient is used to remind us that in evaluating the market solution, we must take into account the costs of acquiring information. Arnott and Stiglitz (1985) have shown that implicit contracts with job turnover are not constrained Pareto efficient because of the externality exerted by firms' wage/employment decisions, via their effect on workers' savings behavior, on other firms' profits; this is an example of the more general class of "seemingly unrelated events" externalities that result whenever there are moral hazard problems (see Arnott and Stiglitz, 1986).

Two related problems should also be mentioned. Firstly, although we have not formally modelled the micro-structure of search, it is clear from the work of Peter Diamond (1982) and Dale Mortensen (1982) that were we to do so, a variety of search externalities would also arise; Hosios (1985) describes the resulting contractual inefficiencies when a Diamond-Mortensen matching model is used to characterize the ex post spot market. Secondly, with risk-averse individuals and more than one commodity, the market equilibrium with limited risk markets will not in general be contrained Pareto efficient; building on work by David Newbery and Stiglitz (1982), Hosios (1984) describes the welfare economics of employment contracts when risk markets are incomplete.

References

Akerloff, George A. and Miyazaki, Hajime, "The Implicit Contract Theory of Unemployment meets the Wage Bill Argument," Review of Economic Studies, January, 1980, 47, 321,338.

Arnott, Richard J. and Stiglitz, Joseph E., "Labor Turnover, Wage Structures and Moral Hazard: The Inefficiency of Competitive Markets", Journal of Labor Economics, October 1985, 3, 434-462.

________ and ________, "The Welfare Economics of Moral Hazard," Institute for Economic Research, Queen's University, Disc. Paper 635, 1986.

________, Hosios, Arthur J. and Stiglitz, Joseph E., "Implicit Contracts, Labor Mobility and Unemployment", Discussion Paper 543, Queen's University, 1983.

Azariadis, Costas, "Implicit Contracts and Underemployment Equilibria", Journal of Political Economy, December 1975, 83, 1183-1202.

________, "Implicit Contracts and Related Topics: A Survey", in The Economics of the Labour Market. Eds. Z. Hornstein et al. London: HMSO, 1979, 221-48.

________, "Employment with Asymmetric Information", Quarterly Journal of Economics (Suppl.), 1983, 98, 157-173.

________ and Stiglitz, Joseph E., "Implicit Contracts and Fixed-Price Equilibria", Quarterly Journal of Economics (Suppl.), 1983, 98, 1-22.

Baily, Martin H., "Wages and Employment under Uncertain Demand", Review of Economic Studies, January 1974, 41, 37-50.

________, "On the Theory of Layoffs and Unemployment", Econometrica, July 1977, 45, 1043-64.

Chan, Ken S. and Ioannides, Yannis M., "Layoff Unemployment, Risk Shifting and Productivity," Quarterly Journal of Economics, May 1982, 97, 213-229.

Chari, V.V., "Involuntary Unemployment and Implicit Contracts," Quarterly Journal of Economics (Suppl.), 1983, 98, 107-123.

Cooper, Russell, "A Note on Overemployment/Underemployment in Labor Contracts under Asymmetric Information," Economic Letters, 1983, 12, 81-89.

Diamond, Peter A., "Aggregate Demand Management in Search Equilibrium," Journal of Political Economy, October 1982, 90, 881-895.

_______________ and Stiglitz, Joseph E., "Increases in Risk and in Risk Aversion", Journal of Economic Theory 1974, 8, 337-60.

Geanakoplos, John and Ito, Takatoshi, "On Implicit Contracts and Involuntary Unemployment," Disc. Paper 81-155R, University of Minnesota,1982.

Green, Jerry and Kahn, Charles, "Wage-Employment Contracts," Quarterly Journal of Economics (Suppl.), 1983, 98, 173-89.

Grossman, Sanford J. and Hart, Oliver D., "Implicit Contracts, Moral Hazard and Unemployment," American Economic Review, May 1981, 301-307.

Hart, Oliver D., "Optimal Labor Contracts under Asymmetric Information: An Introduction," Review of Economic Studies, January 1983, 50, 3-36.

Hosios, Arthur J., "A Welfare Analysis of Employment Contracts with and without Asymmetric Information," Review of Economic Studies, July 1984, 51, 471-489.

_______________, "Excess Capacity, Layoffs and Efficiency," mimeo, University of Toronto, 1985.

_______________, "Layoffs, Recruitment and Interfirm Mobility," Journal of Labor Economics, October 1986, 4, 473-502.

Imai, Haruo, Geanakoplos, John and Ito, Takatoshi, "Incomplete Insurance and Absolute Risk Aversion," Economic Letters, 1981, 8, 107-112.

Ito, Takatoshi, "Labor Contracts with Voluntary Quits", Disc. Paper 233, University of Minnesota, 1986.

Kahn, Charles, M., "Optimal Severance Pay with Incomplete Information", Journal of Political Economy, June 1985, 93, 435-451.

Lowenstein, Mark A., "Worker Heterogeneity, Hours Restrictions and Temporary Layoffs," Econometrica, January 1983, 51, 69-78.

Mortensen, Dale T., "The Matching Process as a Noncooperative Bargaining Game," in The Economics of Information and Uncertainty, edited by John J. McCall, Chicago: University of Chicago Press, 1982.

Newbery, David M.G. and Stiglitz, Joseph E., "The Choice of Technique and the Optimality of Market Equilibrium with Rational Expectations," Journal of Political Economy, 1982, 90, 223-246.

Rosen, Sherwin, "Implicit Contracts: A Survey", Journal of Economic Literature, 1985, 23, 1144-75.

Salop, Steven C., "A Model of the Natural Rate of Unemployment," American Economic Review, March 1979, 69, 117-125.

Stiglitz, Joseph E., "Theories of Wage Rigidity," in Keynes Economic Legacy: Contemporary Economic Theories, edited by James L. Butkiewicz, Kenneth J. Koford and Jeffrey B. Miller, New York: Praeger, 153-206, 1986.

Strand, Jon, "Work Effort and Search Subsidies with Long-Run Equilibrium Contracts", European Economic Review (forthcoming).

Weiss, Andrew, "Job Queues and Layoffs in Labor Markets with Flexible Wages," Journal of Political Economy, June 1980, 88, 526-538.

</ref_section>

PERIOD 1

Workers choose among contracts specifying:

h_L = hours to be worked in the low-productivity state;
h_H = hours to be worked in the high-productivity state;
w_L = payment to employed workers in the low-productivity state;
w_H = payment to employed workers in the high-productivity state;
r = probability of being retained in the low-productivity state;
s = severance payment to laid-off workers.

PERIOD 2

Realization of labor productivity:

$$\theta \in \{\theta_L, \theta_H\} \qquad \theta_L < \theta_H.$$

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 $e(w_i, h_i)$ when $\theta = \theta_i$;
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

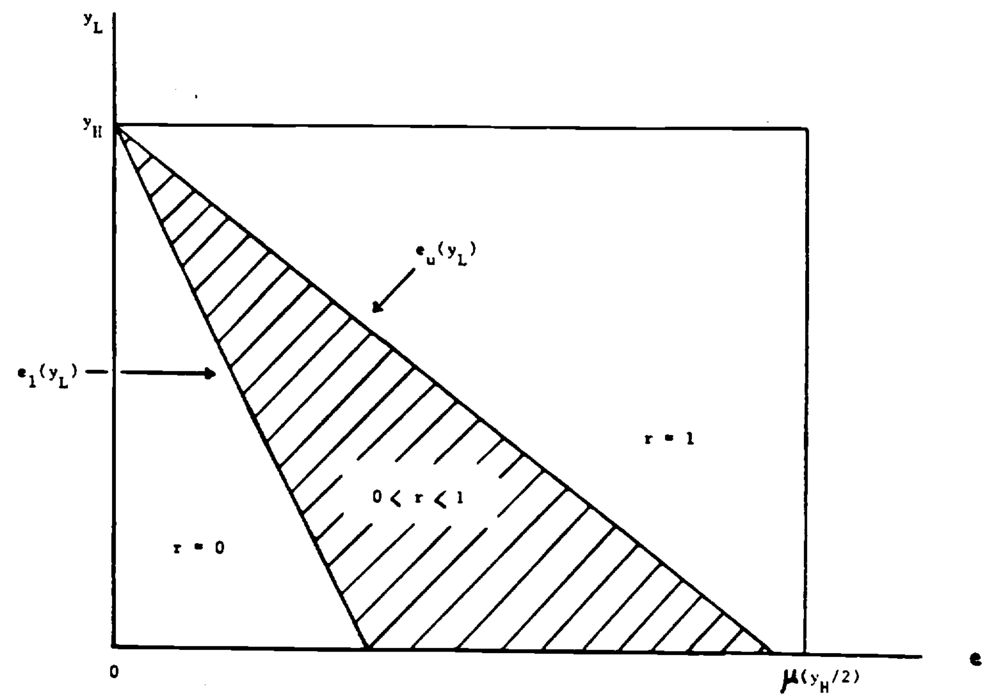

FIGURE 2

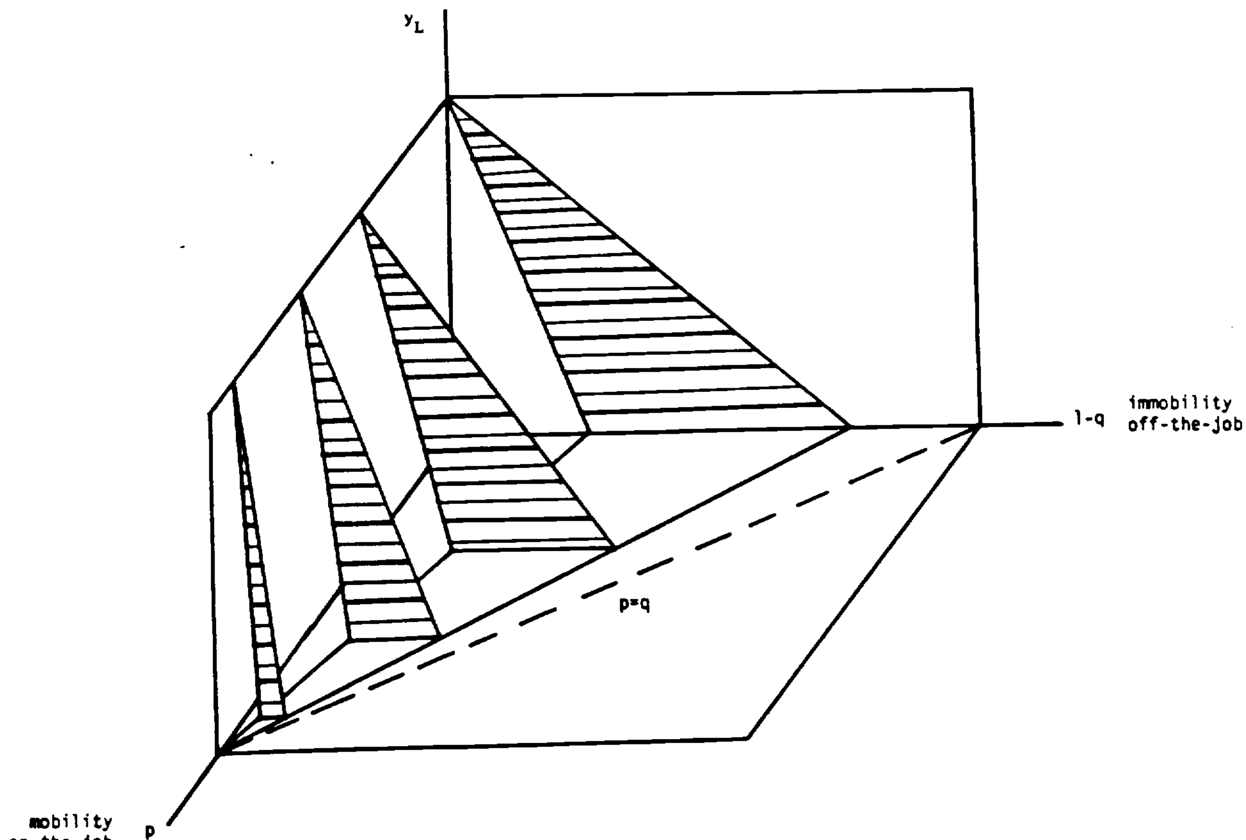

FIGURE 3